STIC-Biotech/ChemLib 79281

From: Ibrahim, Medina A.
Sent: Friday, November 01, 2002 8:29 PM
To: STIC-Biotech/ChemLib
Subject: 09/905, 558

Please search the following:

1. SEQ ID NO:3 and 16. Please search both commercial and issued patents databases. Thanks

Medina A. Ibrahim
Patent Examiner
GAU:1638
CM1-9A12
mailbox9E12
(703)306-5822

RECEIVED
NOV - 4 2002
(STIC)

Mary Jane Ruhl
Tech. Info. Specialist, STIC
TC-1600
CM-1, Room 6A-06
Phone: 605-1155

Mary Jane Ruhl
Tech. Info. Specialist, STIC
TC-1600
CM-1, Room 6A-06
Phone: 605-1155

Searcher:_____
Phone:_____
Location:_____
Date Picked Up:_____
Date Completed:_____
Searcher Prep/Review:_____
Clerical:_____
Online time:_____

TYPE OF SEARCH:
NA Sequences:_____
AA Sequences:_____
Structures:_____
Bibliographic:_____
Litigation:_____
Full text:_____
Patent Family:_____
Other:_____

VENDOR/COST (where applic.)
STN:_____
DIALOG:_____
Questel/Orbit:_____
DRLink:_____
Lexis/Nexis:_____
Sequence Sys.:_____
WWW/Internet:_____
Other (specify):_____

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into to two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**
Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

According to the Pre Publication Rules, every patent application received by the United States Patent and Trademark Office after November 29, 2000 will be pre-published at eighteen months from the effective filing date. When the application is published the contents, including the sequences, will become prior art.

Two new databases have been created to hold the pre-published sequences:
Published_Applications_NA contains nucleic acid sequences; the search results will have the extension **.rnpb**.
Published_Applications_AA contains amino acid sequences; the search results will have the extension **.rapb**.

Each pre-published application is given a unique Publication Number. An example of a Publication Number is US20021234567A1. The "US" indicates the application was a U.S. application. The first 4 digits show the calendar year the application was published. The next 7 digits represent when the application was published. This 7-digit number starts at zero at the beginning of each calendar year. Each application published is given the next number in order. The "A" indicates a utility patent application and the "1" shows that this was the first time the application had been published. If the applicants submit changes to the application, they may requests that the changed application be published again. In such instances, the "1" at the end of the number would be replaced by a "2".

Sequences in the PGPub database are public information; it is permissible to leave these results in the case.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:47:38 ; Search time 1687.14 Seconds
(without alignments)
16248.621 Million cell updates/sec

Title: US-09-905-558C-3
Perfect score: 1310
Sequence: 1 cccatcgctgctttgtctac.....caagggaaagtgatcggatg 1310

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		232.8	17.8	2126	6	AX299951	AX299951 Sequence
2	c	124.6	9.5	1888	6	AX099709	AX099709 Sequence
3	c	112.2	8.6	9285	8	ZMU09989	U09989 Zea mays D3
4		60.2	4.6	174802	2	AC108209	AC108209 Homo sapi
5		58.2	4.4	125020	9	AF429315	AF429315 Homo sapi
6		56.6	4.3	173848	9	AC090043	AC090043 Homo sapi
7		56.2	4.3	169660	9	AL161449	AL161449 Human DNA
8		55.6	4.2	204652	2	PFMAL13P6	AL049183 Plasmodiu
9		55.2	4.2	650	3	SGU36751	U36751 Schizaphis
10	c	54.8	4.2	204652	2	PFMAL13P6	AL049183 Plasmodiu
11		54	4.1	89904	8	ATT25K17	AL049171 Arabidops
12		54	4.1	196286	8	ATCHRIV64	AL161564 Arabidops
13	c	53.8	4.1	115522	8	AC006248	AC006248 Arabidops
14	c	53.6	4.1	214033	2	AC105576	AC105576 Rattus no
15		53.4	4.1	614	3	AF254090	AF254090 Rhopalosi
16	c	53.2	4.1	167553	9	AC090042	AC090042 Homo sapi
17		53	4.0	127539	9	CNS01DSN	AL121821 Human chr
18		53	4.0	146868	2	AC016802	AC016802 Homo sapi
19		53	4.0	151117	2	AL357128	AL357128 Homo sapi
20		53	4.0	161277	2	AC091953	AC091953 Homo sapi
21		53	4.0	176033	2	AL161794	AL161794 Homo sapi
22		53	4.0	178342	9	AC007435	AC007435 Homo sapi
23		53	4.0	187150	2	AC025924	AC025924 Homo sapi
24	c	53	4.0	199385	2	AC019049	AC019049 Homo sapi
25	c	52.6	4.0	72090	2	AC095644	AC095644 Rattus no
26		52.6	4.0	80692	2	AC100569	AC100569 Mus muscu
27		52.6	4.0	110000	2	PFMAL4P1_0	AL034557 Plasmodiu
28		52.6	4.0	191772	2	AC078975	AC078975 Homo sapi
29		52.6	4.0	293431	2	PFMAL13P4	AL049181 Plasmodiu
30		52.4	4.0	171574	2	AC012300	AC012300 Homo sapi
31	c	52.4	4.0	196203	9	AL355355	AL355355 Human DNA
32	c	52.2	4.0	90487	9	AL592166	AL592166 Human DNA
33	c	52.2	4.0	126801	9	AC004849	AC004849 Homo sapi
34	c	52.2	4.0	149450	2	AC067880	AC067880 Homo sapi
35		52.2	4.0	153300	9	AC073387	AC073387 Homo sapi
36	c	52.2	4.0	170141	2	AL611928	AL611928 Homo sapi
37		52.2	4.0	213692	2	AL590427	AL590427 Homo sapi
38	c	52	4.0	159593	9	AC004832	AC004832 Homo sapi
39		51.8	4.0	678	3	AF143595	AF143595 Myzus var
40	c	51.8	4.0	14135	3	AE001388	AE001388 Plasmodiu
41	c	51.6	3.9	79646	8	ATT10D17	AL353865 Arabidops
42		51.6	3.9	162996	9	AC006441	AC006441 Homo sapi
43		51.4	3.9	179307	2	AC104027	AC104027 Homo sapi
44		51.2	3.9	859	8	AF292708	AF292708 Zea diplo
45		51.2	3.9	7218	6	I66494	I66494 Sequence 14

ALIGNMENTS

RESULT 1
AX299951
LOCUS AX299951 2126 bp DNA linear PAT 26-NOV-2001
DEFINITION Sequence 92 from Patent WO0183790.
ACCESSION AX299951
VERSION AX299951.1 GI:17129442
KEYWORDS .
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (sites)
AUTHORS Conner,T.W., Dubois,P., Malven,M. and Masucci,J.D.
TITLE Plant regulatory sequences for selective control of gene expression
JOURNAL Patent: WO 0183790-A 92 08-NOV-2001;
Monsanto Technology LLC (US)
FEATURES Location/Qualifiers
source 1. .2126
/organism="Zea mays"
/db_xref="taxon:4577"
BASE COUNT 607 a 419 c 433 g 667 t
ORIGIN

```
  Query Match             17.8%;  Score 232.8;  DB 6;  Length 2126;
  Best Local Similarity   83.0%;  Pred. No. 1.8e-44;
  Matches  289;  Conservative    0;  Mismatches   57;  Indels    2;  Gaps    2;

Qy       5 TCGCTGCTTTGTCTACATCATGTTCTTCATCATCCTCCCCAGGCGACGCGTGCTGCTGTT 64
Db      82 TCGCGTGATCGTCCACAGCTTGCTGTTCGTCGCCTTCCCAAGTTGACGCGTGCTGCTGTT 141

Qy      65 CTTATTCA-GACTACCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGT 123
Db     142 CTTCTTCCCGGCGACCGTTCGAGGGACTGCACTGCGTACATCTTCCTGCACCGACTTCGT 201

Qy     124 ACGGCTACATCGAACATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAA 183
Db     202 ACGGCTACATCGAACAAACACACGAGATGTCTCGTGTGAATGGAGCCACTGGTGCCTTGA 261

Qy     184 GCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTG 243
Db     262 GCATCGGTCCCTCCGCTGGGTACACTCTGTTCTTCGTATTTGTGCAT-GTTTCATTGCTG 320

Qy     244 TTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACA 303
Db     321 TTTACTGCTTATGCGAGTAGTTATACACATATGCACATACATGTCATCACATATATCGCA 380

Qy     304 ATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACA 351
Db     381 CTGATTATCTGGATTAAATTAAAACTAAAAATGCCTAACTTTCTAACA 428
```

```
RESULT 2
AX099709/c
LOCUS       AX099709                1888 bp    DNA     linear   PAT 02-APR-2001
DEFINITION  Sequence 65 from Patent WO0119976.
ACCESSION   AX099709
VERSION     AX099709.1  GI:13538763
KEYWORDS    .
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1888)
  AUTHORS   Anderson,H.M., Chay,C.A., Chen,G. and Conner,T.W.
  TITLE     Plant regulatory sequences for control of gene expression
  JOURNAL   Patent: WO 0119976-A 65 22-MAR-2001;
            MONSANTO COMPANY (US)
FEATURES             Location/Qualifiers
     source          1. .1888
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
BASE COUNT      569 a    443 c    423 g    453 t
ORIGIN

  Query Match             9.5%;  Score 124.6;  DB 6;  Length 1888;
  Best Local Similarity   63.9%;  Pred. No. 7.1e-19;
  Matches  227;  Conservative    0;  Mismatches  114;  Indels   14;  Gaps    2;

Qy      78 CCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGTACGGCTACATCGAA 137
Db     784 CTGTTTGAGGGACTGCACTGCGTAGACCTGCCTGCATCGAGTTCGTACAATTATATCGAA 725

Qy     138 CATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCC 197
Db     724 CATACACACGAGATGTCTTGTGTGAATTGGGCCGCTGATACTTTGAGCATCGGTCCCTTC 665

Qy     198 GTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATAC 257
Db     664 ACAGGGAACTTCCTGTTTTCCATACTTTT--TGCGTGTCTTTTTTGTGTACTGCTTATAC 607

Qy     258 GAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAAT 317
Db     606 GAATAGTGGT----------ATACATGTCGTCACATAAATTACAATGGTTTTCTGGAC 559
```

```
Qy     318 TAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCA 377
Db     558 TAAATTAAAACTAGTAATGCCTCATTTTCTAACATATTTCACAGTAATAAAAGTTAGGTA 499

Qy     378 ACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACAT 432
Db     498 TCAGCTAATTCTTTGGTGATAGGGTATAAATGACGATGCTCTATATTTATCAGAT 444
```

```
RESULT 3
ZMU09989/c
LOCUS       ZMU09989                9285 bp    DNA     linear   PLN 06-DEC-2000
DEFINITION  Zea mays D3L H(+)-transporting ATPase (Mha1) gene, complete cds.
ACCESSION   U09989
VERSION     U09989.1  GI:507770
KEYWORDS    .
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 2391 to 2453; 2525 to 2659; 2751 to 2930; 3019 to 3225;
            3306 to 3425; 3510 to 3629; 3722 to 3844; 3927 to 3938)
  AUTHORS   Jin,Y. and Bennetzen,J.L.
  TITLE     Integration and Nonrandom Mutation of a Plasma Membrane Proton
            ATPase Gene Fragment within the Bs1 Retroelement of Maize
  JOURNAL   Plant Cell 6, 1177-1186 (1994)
  MEDLINE   95003707
REFERENCE   2  (bases 1 to 9285)
  AUTHORS   Jin,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-1994) Young-Kwan Jin, Biological Sciences, Purdue
            University, 339 Hansen Life Science Research Building, West
            Lafayette, IN 47907, USA
FEATURES             Location/Qualifiers
     source          1. .9285
                     /organism="Zea mays"
                     /strain="D3L"
                     /db_xref="taxon:4577"
                     /chromosome="2L"
                     /map="2L"
                     /clone="C18"
                     /clone_lib="Sau3AI partial of maize genomic DNA into
                     Charon  40"
     mRNA            join(<1446. .1508,1874. .1993,2105. .2203,2316. .2453,
                     2525. .2659,2751. .2930,3019. .3225,3306. .3425,3510. .3629,
                     3722. .3844,3927. .4031,4113. .4257,4379. .4543,4637. .4875,
                     4964. .4996,5091. .5172,5257. .5417,5513. .5689,5768. .5950,
                     6036. .>6290)
                     /gene="Mha1"
                     /product="H(+)-transporting ATPase"
     gene            join(<1446. .1508,1874. .1993,2105. .2203,2316. .2453,
                     2525. .2659,2751. .2930,3019. .3225,3306. .3425,3510. .3629,
                     3722. .3844,3927. .4031,4113. .4257,4379. .4543,4637. .4875,
                     4964. .4996,5091. .5172,5257. .5417,5513. .5689,5768. .5950,
                     6036. .>6290)
                     /gene="Mha1"
     exon            <1446. .1508
                     /gene="Mha1"
                     /number=1
     CDS             join(1446. .1508,1874. .1993,2105. .2203,2316. .2453,
                     2525. .2659,2751. .2930,3019. .3225,3306. .3425,3510. .3629,
                     3722. .3844,3927. .4031,4113. .4257,4379. .4543,4637. .4875,
                     4964. .4996,5091. .5172,5257. .5417,5513. .5689,5768. .5950,
                     6036. .6290)
                     /gene="Mha1"
                     /codon_start=1
                     /product="H(+)-transporting ATPase"
                     /protein_id="AAB60276.1"
                     /db_xref="GI:533775"
                     /translation="MGTTEDKASNLDAVLKEAVDLENIAIQEVFESLRCSPQGLSTEQ
                     AEQRLAIFGPNKLEEKQESKFLKFLGFMWNPLSWVMEAAAIMAIALANGGNKPPDWQD
```

```
                FVGIITLLIINSTISFIEENNAGNAAAALMARLAPKAKVLRNGRWAEEESAILVPGDI
                ISVKLGDIIPADARLLEGDPLKIDQSALTGESLPVTKGPGDGVYSGSTCKQGEIEAVV
                IATGVHTFFGKAAHLVDSTNQVGHFQKVLTAIGNFCICSIAVGMLVEIIVMYPIQHRA
                YRSGIDNLLVLLIGGIPIAMPTVLSVTMAIGAHRLAQQGAITKRMTAIEEMAGMDILC
                SDKTGTLTLNQLTVDKSLVEVFQRGVDQDTVILMAARASRTENQDAIDATIVGMLADP
                TEARAGVQEIHFLPFNPTDKRTALTYLDGEGRMHRVSKGAPEQILHLAHNKKDIETRV
                RAVIDNFAERGLRALGVAYQEVPDGRKESPGGPWEFMGLLPLFDPPRKDSADTISKAL
                DLGVNVKMITGDQLAIGKETARRLGMGTNMYPSSALLEQNKDESIASLPIDELIETAD
                GFAGVFPEHKYEIVKRLQARKHISGMTGDGVNDAPALKKADIGIAVADSTDAARSASD
                IVLTEVGLSVIISAVLTSRAIFQRMKNYTIYAVSITIRIVLGFMLLALIWKFDFPPFM
                VLIIAILNDGTIMTISKDRVKPSPQPDSWKLAEIFATGVVLGTYLAMMTVIFFWAAYK
                TDFFPRLFHVESLAHDDFQMLASAVYLQVSTISQALIFVTRSRSWSFVERPGFLLVSA
                FLVAQLIATLIAVYANWAFAAIKGIGWGWAGVIWLYNIVCYLPLDIIKFLIRYALSGR
                AWNLVLEQRIAFTSKKNFGAEERERRWAHAQRSLHGLELQPPEAASMFENKTSFSEVN
                QLAEEARRRAEMARSDCCLSSISTERGDQTLHCPV"
intron          1509. .1873
                /gene="Mha1"
                /number=1
exon            1874. .1993
                /gene="Mha1"
                /number=2
intron          1994. .2104
                /gene="Mha1"
                /number=2
exon            2105. .2203
                /gene="Mha1"
                /number=3
intron          2204. .2315
                /gene="Mha1"
                /number=3
exon            2316. .2453
                /gene="Mha1"
                /number=4
intron          2454. .2524
                /gene="Mha1"
                /number=4
exon            2525. .2659
                /gene="Mha1"
                /number=5
intron          2560. .2750
                /gene="Mha1"
                /number=5
exon            2751. .2930
                /gene="Mha1"
                /number=6
intron          2931. .3018
                /gene="Mha1"
                /cons_splice=(5'site:no, 3'site:yes)
                /number=6
exon            3019. .3225
                /gene="Mha1"
                /number=7
intron          3226. .3305
                /gene="Mha1"
                /number=7
exon            3306. .3425
                /gene="Mha1"
                /number=8
intron          3426. .3509
                /gene="Mha1"
                /number=8
exon            3510. .3629
                /gene="Mha1"
                /number=9
intron          3630. .3721
                /gene="Mha1"
                /number=9
exon            3722. .3844
                /gene="Mha1"
                /number=10
intron          3845. .3926
                /gene="Mha1"
                /number=10
exon            3927. .4031
                /gene="Mha1"
                /number=11
intron          4032. .4112
                /gene="Mha1"
                /number=11
exon            4113. .4257
                /gene="Mha1"
                /number=12
intron          4258. .4378
                /gene="Mha1"
                /number=12
exon            4379. .4543
                /gene="Mha1"
                /number=13
intron          4544. .4636
                /gene="Mha1"
                /number=13
exon            4637. .4875
                /gene="Mha1"
                /number=14
intron          4876. .4963
                /gene="Mha1"
                /number=14
exon            4964. .4996
                /gene="Mha1"
                /number=15
intron          4997. .5090
                /gene="Mha1"
                /number=15
exon            5091. .5172
                /gene="Mha1"
                /number=16
intron          5173. .5256
                /gene="Mha1"
                /number=16
exon            5257. .5417
                /gene="Mha1"
                /number=17
intron          5418. .5512
                /gene="Mha1"
                /number=17
exon            5513. .5689
                /gene="Mha1"
                /number=18
intron          5690. .5767
                /gene="Mha1"
                /number=18
exon            5768. .5950
                /gene="Mha1"
                /number=19
intron          5951. .6035
                /gene="Mha1"
                /number=19
exon            6036. .one-of(6408,6409)
                /number=20
                /evidence=experimental
polyA_site      6408
                /evidence=experimental
polyA_site      6409
                /evidence=experimental
BASE COUNT     2329 a   2181 c   2086 g   2689 t
ORIGIN

  Query Match             8.6%;  Score 112.2;  DB 8;  Length 9285;
  Best Local Similarity   74.1%;  Pred. No. 8.8e-16;
  Matches  177;  Conservative    0;  Mismatches   43;  Indels   19;  Gaps    2;

Qy      78 CCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGTACGGCTACATCGAA 137
           ||||||||| |||||||  | ||||| || |||||||||| |||||| | ||||||||||
Db     220 CCGTTCGAGGGACTGCACTGTGTACACCTGTCTGCATCGAATTTGTATGACTACATCGAA 161

Qy     138 CATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCC 197
```

```
          |||| |||||||||||||| ||||||| ||| |||| |||                    |
Db    160 CATACACACGAGATGTCTCATGTGAATGGAGCCACTGATG------------------TC 119

Qy    198 GTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATAC 257
          | ||||||| ||    | | | || |||||||  ||| |||  ||||||||  |||| |
Db    118 GCAGGGTACTTTTAAATTTCCATACTTGTGCAT-GTTTCATTTATGTTTACTACTTATGC 60

Qy    258 GAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAA 316
          ||||||  |||| |||| ||||||||| ||||||||||||||||  | || ||||| |
Db     59 CAGTAGTGGTACACACATACACATACATGTCATCACATATATCACGTTGTTCTTCTAGA 1


RESULT 4
AC108209
LOCUS       AC108209              174802 bp    DNA     linear   HTG 31-JAN-2002
DEFINITION  Homo sapiens chromosome 4 clone RP11-439B3, WORKING DRAFT SEQUENCE,
            7 unordered pieces.
ACCESSION   AC108209
VERSION     AC108209.2  GI:18450201
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174802)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174802)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jan 31, 2002 this sequence version replaced gi:18376960.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0439B03
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 2%
            Sequencing vector: plasmid; 98%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 173471 bases at least Q40
            Consensus quality: 173768 bases at least Q30
            Consensus quality: 173922 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 174202; sum-of-contigs
            Quality coverage: 12.07 in Q20 bases; agarose-fp
            Quality coverage: 12.00 in Q20 bases; sum-of-contigs
            ------------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1162: contig of 1162 bp in length
            *     1163     1262: gap of unknown length
            *     1263     2346: contig of 1084 bp in length
            *     2347     2446: gap of unknown length
            *     2447     8520: contig of 6074 bp in length
            *     8521     8620: gap of unknown length
            *     8621    24285: contig of 15665 bp in length
            *    24286    24385: gap of unknown length
            *    24386    63017: contig of 38632 bp in length
            *    63018    63117: gap of unknown length
            *    63118   113781: contig of 50664 bp in length
            *   113782   113881: gap of unknown length
            *   113882   174802: contig of 60921 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174802
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-439B3"
     misc_feature    1. .1162
                     /note="assembly_name:Contig23"
     misc_feature    1263. .2346
                     /note="assembly_name:Contig27"
     misc_feature    2447. .8520
                     /note="assembly_name:Contig32"
     misc_feature    8621. .24285
                     /note="assembly_name:Contig33"
     misc_feature    24386. .63017
                     /note="assembly_name:Contig34
                     clone_end:T7
                     vector_side:left"
     misc_feature    63118. .113781
                     /note="assembly_name:Contig35
                     clone_end:SP6
                     vector_side:right"
     misc_feature    113882. .174802
                     /note="assembly_name:Contig36"
BASE COUNT    54963 a  33646 c  34740 g  50806 t    647 others
ORIGIN

  Query Match             4.6%;  Score 60.2;  DB 2;  Length 174802;
  Best Local Similarity   52.6%;  Pred. No. 0.0035;
  Matches  131;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy    204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
          || | | | ||  | | |||     |   | |  || |   | | | | |||     ||
Db  58325 TATAATATATTTATATAATTATATAAATGTATACTTATATATAATTCAATATTTATATAA 58384

Qy    264 TTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATT 323
          ||||| |||  |  | ||  |||| || | |||||   |||||||| | ||| |  |  |
Db  58385 TTATATATAATTCAATATTTATATAATTATATATAATTCAATATTTATATAATTATATAT 58444

Qy    324 AAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACT 383
          ||  | | |  |   ||| | | ||        | | |||   ||  | | |  |||| |
Db  58445 AATTCAATATTTATATAATTATATATATAATTCAATATTTATATAATATATTTAACATAT 58504

Qy    384 TTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATG 443
          |  |||  |   | || ||| |    |||   | ||||||  |   ||||| ||
Db  58505 ATTTATATAATTATATATTATATATAATTTTTGTATATATAAAATATTCAGTCTTAAAAA 58564

Qy    444 GAAAACAGT 452
          ||| ||| |
Db  58565 GAAGACAAT 58573


RESULT 5
AF429315
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
```

```
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN

  Query Match             4.4%;  Score 58.2;  DB 9;  Length 125020;
  Best Local Similarity   10.6%;  Pred. No. 0.0096;
  Matches   60;  Conservative  248;  Mismatches  256;  Indels    0;  Gaps    0;

Qy     745 GCACGAAGGGCACTACCCCAACCTCTCACCGAAAAACCGCGCTGGATCGGCAAATCAAAC 804
           ::::::  ::: :  :::::|:: | :::  : :  |::: :::::: ::  :    ::
Db   16892 KSWMSMBMSVSYSVKMHSMHASBSCMHWBKCMTWSCCMSMMYKSSWWGSSWGMCCWGRRR 16951

Qy     805 GAGGTGGTGCCCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCAC 864
           : | :    : :: :::: :   :::::::: :| :: :|::  :||::  |:::| | |::
Db   16952 SKGWKWYSRGMSRSKSMRYTGGSKMMRSSMMCTSSCYASMCCMCCWSCCCMRSCCCMCRS 17011

Qy     865 CAGCCATGCCGTGTCGCGGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCC 924
           :  |::: |  : ::::   :: ::::::::::  ::  :: : : ::::   :  || ::
Db   17012 YCCCMRYCCACKCYMSSYWTMSASYMSRRSYWKRSMKCWSRMCSRSRSSKCKSRGCGGSM 17071

Qy     925 CGGCCGTGCCCGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCT 984
           :|  |:::  :|:  ::  ::   ::  :::: ::    :: |: |  |: :|   |:
Db   17072 GKGKGGKSYGRKTKRKSRGKMKGAKWYMYYRRSMRKMKMYSSKGMYCMYCWCWGRRGCYC 17131

Qy     985 GGACACCCGACACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCAC 1044
           : :   :   | :|: :  : |:: ::|     ::  |:::  : :::    : :: :
Db   17132 SCMTSRSAMCCSYCYAKCKSMCYSCYYGMSMKGYYYACSYRGSMSSKYCMRGSTYSTSCG 17191

Qy    1045 GGCCGACTGCGCTGCGCCGCCCGGATATATATACCCATCGTTATCGATCGATCGATCGCG 1104
             |   | | |      |   | | |       : ::::  | :     :   :  :
Db   17192 CCCTTTTTCCCCCNANTGGGGAAGCTTTTNCNKKTYSYYRKRNGCAMCKYNNYYNNSWRS 17251

Qy    1105 TCACTCACGGGTAGCTCATGGTCGAGCGTAGCATGCAGGAACTTATTTGCCGTGCGCTCC 1164
           |:  : | : : ::|:   ::|   : :  :: ::|:: :::  : : :    : ::
Db   17252 SCRAGMSCTKYKSSMTMSMASYCWCMSMYYCSMRSMASRGMSWSSYMKMMKSSWRMSCYM 17311

Qy    1165 CAGGTCTCCGCTCGCGTGCCTTCCAGTCTGTCTCACACTAGCTGCTGTGGGACGATCGAA 1224
             :::| | :::  :    :::  ::: :   | ||: :: : :::   :   : :
Db   17312 KCCWCMKMCYCMMRSMRSRGMSYYMYASWKSSSRGCYCTRCYWCMSSKSCYKSYYMMMRS 17371

Qy    1225 GTGGGTGTGTCAGCTAGCTAGCTGCGCCGTGACCACGCACATGACCGCAGTGCGCGCGGG 1284
           :  :| ::: |:|: :|:  ::  ::|::  :|:::: |   | | ::: |:   ::
Db   17372 KRMKGMKWSRCWGSMWGSASRSSCYKCYKSMRCSMMSSKCYRCAGCMMMKGGYMRYMRCW 17431

Qy    1285 GCTGATCAAGGGAAAGTGATCGGA 1308
           ::::::   :   :::  :: :||
Db   17432 SMKRRWWGKSAMYRMRRMWKGRGA 17455


RESULT 6
AC090043
LOCUS       AC090043              173848 bp    DNA     linear   PRI 11-FEB-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-551L4 map 3p, complete
            sequence.
ACCESSION   AC090043
VERSION     AC090043.1  GI:12745081
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173848)
  AUTHORS   Xiong,H., Zhou,Y., Dong,H., Lin,W., Chen,B., Zhang,C., Zhang,Y.,
            Cai,Z., Ying,H.F., Wang,H., Gu,W., Zhu,G., Tu,Y., Zhang,X., Jia,J.,
            Shen,H., Zhang,D., Wu,C., Lu,G., Zhong,M., Jiang,H., Ren,S., Fu,G.,
            Chen,Z. and Huang,M.
  TITLE     Chromosome 3p genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173848)
  AUTHORS   Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,C., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
            Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X.,
            Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2001) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
COMMENT     --------------Genome Center
            Center:Beijing Center
            Center code:Beijing
            Website:http://hgc.igtp.ac.cn
            http://www.genomics.org.cn
            Contact:hgc@igtp.ac.cn
            -------------- Project Information
            Center project name:1% project
            Center clone name: RP11-551L4
            -------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator: ET 55% of reads
            Chemistry: Dye-terminator Big Dye; 45% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 501 bases at least Q40
            Consensus quality: 829 bases at least Q30
            Consensus quality: 1046 bases at least Q20
            Insert size: 1157; sum-of-contigs
            Quality coverage:   1.35x in Q20 bases;sum-of-contigs
            --------------.
FEATURES             Location/Qualifiers
     source          1. .173848
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
```

```
                     /map="3p"
                     /clone="RP11-551L4"
BASE COUNT    56663 a  31451 c  31882 g  53852 t
ORIGIN

  Query Match             4.3%;  Score 56.6;  DB 9;  Length 173848;
  Best Local Similarity   47.0%;  Pred. No. 0.025;
  Matches  208;  Conservative    0;  Mismatches  234;  Indels    1;  Gaps    1;

Qy    100 TACATCTTTCTGCATCGACTTTGTACGGCTACATCGAACATATACACGAGATGTCTCGTG 159
          || |  ||| |  ||  | | | || |  |||||  |    |        | || | | ||
Db 163709 TATAAATTTATATATATATTATATATGTATACATATATATTTATATATAAATTTATAGTT 163768

Qy    160 TGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCT 219
          |  ||| |   |   ||   |          |||     |   ||  |    ||
Db 163769 TACATATAAATAGATATATTTATTTATATATTTATATATTTATATATTTATATTTTAATA 163828

Qy    220 TATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 279
          | | | |  |||||| ||| ||| | ||   |||||| |  ||  |||| ||||||  |
Db 163829 TGTATTTATATATTTATATAGTTATATATAAATTATATGTATAAATATATATACATATAT 163888

Qy    280 ATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACT 339
          ||  ||||    ||||||||    |||||| |  | ||  |||  |   || |  |   |
Db 163889 ATGGATATTTATACATATATATTTATATTTATACATATATAATATATTATATATTTATAT 163948

Qy    340 AAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTAT-TCTACAT 398
            ||   ||      |     || || || |   |  |     | ||  |||   |||||
Db 163949 TTATACATATGTATGATATATTTATATTGATACATATATATTATATATTTATAAATACAT 164008

Qy    399 TGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTG 458
          | ||   ||     | | | | | || | |      | ||||     ||| |   | |
Db 164009 TATTTATTTATATGTATAAATATTTATATATGTATATATACATATATAAATATATGTATA 164068

Qy    459 TACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAGTCGTTTGAAGATTGA 518
          |     ||||      || |||   || |     | |  |||||  | | |  | |||
Db 164069 TGTACATATAATTTATATATGTACATATACATATATAATATATATACATATATAAATTAT 164128

Qy    519 ACCTATATATCGGTACGGTTAAT 541
          |  || ||||    |    ||||
Db 164129 ATGTACATATACACATATATAAT 164151


RESULT 7
AL161449
LOCUS       AL161449               169660 bp    DNA     linear   PRI 04-DEC-2001
DEFINITION  Human DNA sequence from clone RP11-382H24 on chromosome 9p22.1-23
            Contains a thioredoxin peroxidase pseudogene, a SSB (Sjogren
            syndrome antigen B (autoantigen La) pseudogene, the 3' end of the
            MPDZ gene for multiple PDZ domain protein and a CpG island,
            complete sequence.
ACCESSION   AL161449
VERSION     AL161449.7  GI:8894259
KEYWORDS    HTG; CpG island; MPDZ.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169660)
  AUTHORS   Sehra,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 1, 2000 this sequence version replaced gi:8653807.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            This sequence is the entire insert of clone RP11-382H24 The true
            left end of clone RP11-664D14 is at 135950 in this sequence. The
            true right end of clone RP11-187K14 is at 53851 in this sequence.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP11-382H24 is from
            the library RPCI-11.2 constructed by the group of Pieter de Jong.
            For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .169660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="p22.1-23"
                     /clone="RP11-382H24"
                     /clone_lib="RPCI-11.2"
     misc_feature    2. .492
                     /note="match: GSS: Em:AQ700717"
     misc_feature    complement(1479. .2009)
                     /note="match: GSS: Em:AQ569033"
     repeat_region   1589. .1778
                     /note="MIR repeat: matches 23. .192 of consensus"
     repeat_region   1669. .1724
                     /note="2 copies 28 mer 92% conserved"
     repeat_region   2068. .2401
                     /note="L1MC4 repeat: matches 7453. .7834 of consensus"
     repeat_region   2409. .2828
                     /note="15 copies 28 mer 62% conserved"
     repeat_region   2429. .2848
                     /note="7 copies 60 mer 65% conserved"
     repeat_region   2452. .2801
                     /note="175 copies 2 mer ta 61% conserved"
     repeat_region   2819. .2911
                     /note="L1MC4 repeat: matches 7864. .7961 of consensus"
     misc_feature    complement(4102. .4478)
                     /note="match: GSS: Em:AQ180899"
     misc_feature    4497. .5002
                     /note="match: GSS: Em:AQ716496"
     repeat_region   4621. .4865
                     /note="MIR repeat: matches 1. .261 of consensus"
     repeat_region   5200. .5294
                     /note="L2 repeat: matches 2578. .2670 of consensus"
     repeat_region   6031. .6654
                     /note="L1PA12 repeat: matches 5583. .6152 of consensus"
     misc_feature    7419. .7966
                     /note="match: GSS: Em:AQ378047"
     repeat_region   7431. .7737
                     /note="AluSc repeat: matches 1. .307 of consensus"
     repeat_region   7763. .8103
                     /note="MLT1A1 repeat: matches 205. .536 of consensus"
     repeat_region   8379. .8422
                     /note="22 copies 2 mer tg 86% conserved"
     repeat_region   9171. .9447
                     /note="L1M4 repeat: matches 2900. .3176 of consensus"
     repeat_region   9450. .9739
                     /note="L1M4 repeat: matches 3869. .4175 of consensus"
     repeat_region   9774. .9835
                     /note="L1MA7 repeat: matches 6223. .6284 of consensus
                     L1MA7 repeat: matches 6223. .6284 of consensus"
     repeat_region   10099. .10533
                     /note="MLT1C repeat: matches 9. .466 of consensus"
```

```
repeat_region   10789. .11105
                /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region   11109. .11267
                /note="MER5B repeat: matches 1. .178 of consensus"
misc_feature    complement(12290. .12569)
                /note="match: STS: Em:G05562"
gene            complement(12572. .13167)
                /gene="bA382H24.1"
                /pseudo
CDS             complement(12572. .13167)
                /gene="bA382H24.1"
                /note="bA382H24.1 (thioredoxin peroxidase pseudogene)
                match: cDNAs: Em:U27125 Em:AF221841
                match: proteins: Sw:Q91191 Tr:O93241 Sw:Q06830 Sw:P32119
                Sw:Q63716 Sw:P35704 Sw:Q90384 Tr:O44366 Sw:P35700
                Sw:Q61171 Tr:O96763 Tr:O61000 Tr:O85761"
                /codon_start=1
                /pseudo
                /evidence=not_experimental
misc_feature    14831. .15165
                /note="match: STS: Em:HS260YA5"
repeat_region   14993. .15030
                /note="19 copies 2 mer ac 97% conserved"
repeat_region   16682. .16772
                /note="MLT2D repeat: matches 1. .90 of consensus"
repeat_region   16759. .16999
                /note="MLT1A1 repeat: matches 1. .241 of consensus"
repeat_region   17339. .17638
                /note="AluY repeat: matches 4. .304 of consensus"
repeat_region   17645. .18235
                /note="L2 repeat: matches 1961. .2750 of consensus"
repeat_region   18526. .18676
                /note="MLT1J repeat: matches 33. .189 of consensus"
repeat_region   18882. .18943
                /note="MLT1J repeat: matches 411. .467 of consensus"
repeat_region   18986. .19291
                /note="AluSx repeat: matches 1. .307 of consensus"
repeat_region   19450. .19635
                /note="MIR repeat: matches 59. .248 of consensus"
repeat_region   20230. .20295
                /note="33 copies 2 mer at 75% conserved"
misc_feature    complement(21475. .21909)
                /note="match: GSS: Em:B46263"
repeat_region   21640. .22131
                /note="MLT1H repeat: matches 40. .547 of consensus"
repeat_region   22141. .22256
                /note="MLT1H repeat: matches 236. .352 of consensus"
repeat_region   22708. .22940
                /note="L1PA2 repeat: matches 5403. .5632 of consensus
                L1PA3 repeat: matches 5401. .5632 of consensus"
repeat_region   22941. .23434
                /note="L1PA2 repeat: matches 5641. .6144 of consensus"
repeat_region   22941. .23436
                /note="L1PA3 repeat: matches 5641. .6146 of consensus"
misc_feature    24368. .24949
                /note="match: GSS: Em:AQ744132"
misc_feature    24381. .24772
                /note="match: GSS: Em:AQ333224"
repeat_region   26193. .26290
                /note="MIR repeat: matches 97. .193 of consensus"
misc_feature    complement(26637. .27057)
                /note="match: GSS: Em:B75037"
repeat_region   27837. .27899
                /note="L2 repeat: matches 2641. .2708 of consensus"
repeat_region   28653. .28739
                /note="AluJo/FRAM repeat: matches 202. .289 of consensus"
repeat_region   29135. .29237
                /note="MIR repeat: matches 97. .194 of consensus"
repeat_region   29262. .29957
                /note="L2 repeat: matches 1986. .2742 of consensus"
repeat_region   29958. .30391
                /note="MLT2B repeat: matches 1. .444 of consensus
                MLT2B repeat: matches 1. .444 of consensus"
repeat_region   30392. .30590
                /note="L2 repeat: matches 1771. .1986 of consensus
                HERVL repeat: matches 5456. .5654 of consensus"
repeat_region   30591. .30631
                /note="MLT2B repeat: matches 1. .41 of consensus
                MLT2A repeat: matches 1. .41 of consensus"
repeat_region   30632. .31387
                /note="L2 repeat: matches 1156. .1771 of consensus"
repeat_region   31401. .31876
                /note="MLT1H repeat: matches 13. .516 of consensus"
repeat_region   31936. .32142
                /note="L2 repeat: matches 929. .1149 of consensus"
repeat_region   32790. .33087
                /note="AluSx repeat: matches 1. .297 of consensus"
misc_feature    complement(34142. .34617)
                /note="match: GSS: Em:AQ799891"
misc_feature    34145. .34705
                /note="match: GSS: Em:AQ350618
                match: STS: Em:G57679"
repeat_region   35089. .35217
                /note="L1 repeat: matches 4928. .5057 of consensus"
repeat_region   35239. .36006
                /note="L1ME1 repeat: matches 5352. .6103 of consensus"
repeat_region   37784. .37985
                /note="L1MC4 repeat: matches 5695. .5897 of consensus"
```

```
Query Match             4.3%;  Score 56.2;  DB 9;  Length 169660;
Best Local Similarity   50.2%;  Pred. No. 0.031;
Matches  139;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy     245 TTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACAA 304
           |||   ||||||  | ||  |||   || ||  | | | |||| || | |||    | |
Db    2444 TTATATATTATATAATTATATATTATTAAATATATAAATATATAATTATATAATATATAT 2503

Qy     305 TATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGT 364
           | | |   || ||   ||||||  || ||||   ||| | | |||     | | || |
Db    2504 TTTATAATTATATATTATTAAATATATAAATATATAATTATATAATTTATATTATATAAT 2563

Qy     365 AATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATA 424
           || | || |  ||  |  |    |||  || ||  ||     |  |   |   |  ||
Db    2564 TATATATTATATAATTATATATTATATAATATATAATTTATAATATATAATATATAATTA 2623

Qy     425 AACAACATAGTCTACAATGGAAAACAGTGCTTTGTACGACTATATACGCGATGTGTGGCT 484
           | || |||   |   ||  || | |    | | ||    ||||||     | | |   |
Db    2624 TATAATATATAATTTTATAAAATATAAATATGTTTAAATTTATATATAAATTATATTTAT 2683

Qy     485 ACAACATAAGACAATATAGTCGTTTGAAGATTGAACC 521
           | ||  | | | ||||||    |||  | | | || |
Db    2684 ATAATTTTATAAAATATAAATTTTTATATAATAAAAC 2720
```

```
RESULT 8
PFMAL13P6
LOCUS       PFMAL13P6             204652 bp    DNA     linear   HTG 19-AUG-1999
DEFINITION  Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL049183
VERSION     AL049183.5  GI:5763804
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 204652)
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Aug 24, 1999 this sequence version replaced gi:5731886.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
```

```
          correct sequence. Work on the sequence is in progress and the
          release of this data is based on the understanding that the
          sequence may change as work continues. The sequence may be
          contaminated with foreign sequence from E.coli, yeast, vector,
          phage etc.
          Order of segments is not known; 800 n's separate segments.
          * NOTE: This is a 'working draft' sequence.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
FEATURES             Location/Qualifiers
     source          1. .204652
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="13"
BASE COUNT    68880 a  17396 c  17710 g  67856 t  32810 others
ORIGIN

  Query Match             4.2%;  Score 55.6;  DB 2;  Length 204652;
  Best Local Similarity   52.6%;  Pred. No. 0.044;
  Matches  121;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy     204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
           | |||   || |   || | | |  |||| | |||   | |      ||||||  | ||
Db   27538 TTCATGTGGTACAATTTGTATATATATATGTATATATATATAACAATATTATATAATTAT 27597

Qy     264 TTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATT 323
           |||||||| |  |||||| ||    || |||||| |  |||| ||  ||     ||||
Db   27598 ATATACATATACACACATACGTACATACATATATATAAATATATATTATTACTAACAATT 27657

Qy     324 AAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACT 383
           |||| |||||  |  | ||   ||   || |    | | ||    | | |  |
Db   27658 TAAACAAAAAAATAAAATATAATTATTCCCCAAATAAATATATGAATGTATATATATTTA 27717

Qy     384 TTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATA 433
           |  |  | | ||  || |  ||| |  |  |  | |   ||   ||||||
Db   27718 TGTCTAAATATAGTTTTTCATATATATATATATATTACTTATTAAACATA 27767


RESULT 9
SGU36751
LOCUS       SGU36751                650 bp    DNA     linear   INV 06-FEB-1997
DEFINITION  Schizaphis graminum cytochrome oxidase subunit II (COII) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U36751
VERSION     U36751.1  GI:1036830
KEYWORDS    .
SOURCE      greenbug.
  ORGANISM  Mitochondrion Schizaphis graminum
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha;
            Aphidiformes; Aphidoidea; Aphididae; Aphidini; Schizaphis.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Rouhbakhsh,D., Lai,C.Y., von Dohlen,C.D., Clark,M.A., Baumann,L.,
            Baumann,P., Moran,N.A. and Voegtlin,D.J.
  TITLE     The tryptophan biosynthetic pathway of aphid endosymbionts
            (Buchnera): genetics and evolution of plasmid-associated
            anthranilate synthase (trpEG) within the aphididae
  JOURNAL   J. Mol. Evol. 42 (4), 414-421 (1996)
  MEDLINE   96215866
REFERENCE   2  (bases 1 to 650)
  AUTHORS   von Dohlen,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-1995) Biological Sciences, Idaho State
            University, Campus Box 8007, Pocatello, ID 83209, USA
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Schizaphis graminum"
                     /organelle="mitochondrion"
                     /db_xref="taxon:13262"
     gene            1. .650
                     /gene="COII"
     CDS             <1. .>650
                     /gene="COII"
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome oxidase subunit II"
                     /protein_id="AAB42168.1"
                     /db_xref="GI:1036831"
                     /translation="MTWLKLSFQNSNSPLMEQLIFFHDHTIFIIIMIMTMITYMMIFT
                     MKNKFINIKISENQMIEFIWTTTPPIILIFIAMPTLHLSYSMDEINCPILKIKIFGHQ
                     WFWSYEYSDFSNIEFESYMMNELNKENFRLIVVDNKTIIPFKFNIRLLISSDDVIHSW
                     TIPSLAIKIDSIPGRMNQINLFMNRPGMYFGQCSEICGINHSFMPIQIESINLNKF"
BASE COUNT      265 a     82 c     49 g    254 t
ORIGIN

  Query Match             4.2%;  Score 55.2;  DB 3;  Length 650;
  Best Local Similarity   56.4%;  Pred. No. 0.015;
  Matches  123;  Conservative    0;  Mismatches   93;  Indels    2;  Gaps    1;

Qy     235 TTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACA 294
           ||| | || |||  |||| |||| | |   |  ||     ||    |||   || || ||
Db      55 TTAATTTTTTTTCATGATCATACAATTTTTATTATTATTATAATTATAACAATAATTACT 114

Qy     295 TATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCA 354
           ||||| | |  ||||||   |   ||| ||||  ||  |||    |||| ||| | | |
Db     115 TATATAATA--ATTTTTACTATAAAAAATAAATTTATTAATATTAAAATCTCTGAAAACC 172

Qy     355 ACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATT 414
           |    |||| | |  |||   | ||||||   || ||| | | |  || |||| | |
Db     173 AAATAATTGAATTTATTTGAACTACAACTCCTCCAATTATTCTTATTTTTATTGCTATAC 232

Qy     415 TCACTCTATAAACAACATAGTCTACAATGGAAAACAGT 452
           ||||||  |     |||| || | |   ||||  | |
Db     233 CTACTCTACATCTTTCATATTCAATAGATGAAATTAAT 270


RESULT 10
PFMAL13P6/c
LOCUS       PFMAL13P6               204652 bp    DNA     linear   HTG 19-AUG-1999
DEFINITION  Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL049183
VERSION     AL049183.5  GI:5763804
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 204652)
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Aug 24, 1999 this sequence version replaced gi:5731886.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            Order of segments is not known; 800 n's separate segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .204652
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="13"
```

```
BASE COUNT    68880 a  17396 c  17710 g  67856 t  32810 others
ORIGIN

  Query Match             4.2%;  Score 54.8;  DB 2;  Length 204652;
  Best Local Similarity   51.3%;  Pred. No. 0.069;
  Matches  153;  Conservative    0;  Mismatches  142;  Indels    3;  Gaps    1;

Qy     173 TAATGCCTTAAGCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATAT 232
           ||||     ||  ||    || ||   |   || |||     | | ||| || | ||| |
Db  154818 TAATAATAAAAATATATTATATTCTTAATAATAAATTGAACACATTTTAATTTTACATTT 154759

Qy     233 TTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCA 292
           |      | | |  |   ||||  |  || |||| ||| ||  | ||| ||||    |
Db  154758 GTAATACAATATATTTTTGATATATCATAAGATATATATATATATATATATATATATA 154699

Qy     293 CATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACAC 352
           ||||||    |||| | |   |  || ||  ||  ||   ||   || |||| | | |
Db  154698 TATATATATATATATATATAATATATATATATAATGTAGTTATTTATATATTTATTATAT 154639

Qy     353 CAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAA 412
           | |||   ||   ||||| |   | |     |||   | |||| | ||||   ||
Db  154638 ATTATATATTTGTATAGATTCTCAATTTAATGCCTATATGAAAAATTGATATATT---AA 154582

Qy     413 TTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGTACGACTATATA 470
           ||||  | || |||||| |||    | ||  | ||  | |  | |  |    || |||
Db  154581 TTTCCATATAAAAACAATATATATAAAAAAAGTAATAAATAATATAAATATATAAATA 154524


RESULT 11
ATT25K17
LOCUS       ATT25K17               89904 bp    DNA     linear   PLN 27-AUG-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone  (ESSA project).
ACCESSION   AL049171
VERSION     AL049171.1  GI:4539415
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 89904)
  AUTHORS   Bevan,M., Koetter,P., Hempel,S., Entian,K.-D., Bancroft,I.,
            Mewes,H.W., Mayer,K.F.X. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89904)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosome 4 can be
            viewed at: http://websvr.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .89904
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     misc_feature    1. .11070
                     /note="overlap to BAC ; please refer to this entry for
                     analysis and annotation"
     exon            19417. .19605
                     /gene="T25K17.10"
                     /number=1
     gene            19417. .21056
                     /gene="T25K17.10"
     CDS             join(19417. .19605,19706. .20001,20198. .21056)
                     /gene="T25K17.10"
                     /EC_number="4.4.1.14"
                     /note="strong similarity to ACC synthase , Malus
                     domestica, U73816
                     Contains Aminotransferases class-I pyridoxal-phosphate
                     attachmentsite, Aa_Transfer_Class_1 [SLSKDLGLPGFRVG]
                     contains EST gb:N65643, AA650853, Z33841"
                     /codon_start=1
                     /product="1-aminocyclopropane-1-carboxylate synthase-like
                     protein"
                     /protein_id="CAB38949.1"
                     /db_xref="GI:4539416"
                     /translation="MGLPLMMERSSNNNNVELSRVAVSDTHGEDSPYFAGWKAYDENP
                     YDESHNPSGVIQMGLAENQVSFDLLETYLEKKNPEGSMWGSKGAPGFRENALFQDYHG
                     LKTFRQAMASFMEQIRGGKARFDPDRIVLTAGATAANELLTFILADPNDALLVPTPYY
                     PGFDRDLRWRTGVKIVPIHCDSSNHFQITPEALESAYQTARDANIRVRGVLITNPSNP
                     LGATVQKKVLEDLLDFCVRKNIHLVSDEIYSGSVFHASEFTSVAEIVENIDDVSVKER
                     VHIVYSLSKDLGLPGFRVGTIYSYNDNVVRTARRMSSFTLVSSQTQHMLASMLSDEEF
                     TEKYIRINRERLRRRYDTIVEGLKKAGIECLKGNAGLFCWMNLGFLLEKKTKDGELQL
                     WDVILKELNLNISPGSSCHCSEVGWFRVCFANMSENTLEIALKRIHEFMDRRRRF"
     intron          19606. .19705
                     /gene="T25K17.10"
                     /number=1
     exon            19706. .20001
                     /gene="T25K17.10"
                     /number=2
     intron          20002. .20197
                     /gene="T25K17.10"
                     /number=2
     exon            20198. .21056
                     /gene="T25K17.10"
                     /number=3
     gene            26480. .27228
                     /gene="T25K17.20"
     CDS             join(26480. .26661,27042. .27228)
                     /gene="T25K17.20"
                     /note="strong similarity to predicted protein, Arabidopsis
                     thaliana, AC005169"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB38950.1"
                     /db_xref="GI:4539417"
                     /translation="MASKLLQLKSKACEASKFVSKHGTTYYKQLLDKNKMYIQEPATI
                     EKCNELSKQLLYTRLASIPGRSESFWKEVNHVKGLWKNRADLKVEDAGIAALFGLECF
                     AWYCAGEIVGRGFTFTGYYP"
     exon            26480. .26661
                     /gene="T25K17.20"
                     /number=1
     intron          26662. .27041
                     /gene="T25K17.20"
                     /number=1
     exon            27042. .27228
                     /gene="T25K17.20"
                     /number=2
     exon            28289. .28369
                     /gene="T25K17.30"
                     /number=1
     gene            28289. .29256
                     /gene="T25K17.30"
     CDS             join(28289. .28369,28417. .28641,28748. .28879,28966. .29256)
                     /gene="T25K17.30"
                     /note="strong similarity to caffeoyl-CoA
                     O-methyltransferase, Pinus taeda, AF036095"
                     /codon_start=1
                     /product="caffeoyl-CoA O-methyltransferase-like protein"
                     /protein_id="CAB38951.1"
                     /db_xref="GI:4539418"
                     /translation="MAKDEAKGLLKSEELYKVCPYILISTLYILETSVYPREPEVLRE
                     LRNITHNHPQAGMATAPDAGQLMGMLLNLVNARKTIEVGVFTGYSLLLTALTLPEDGK
                     VIAIDMNRDSYEIGLPVIKKAGVEHKIDFKESEALPALDELLNNKVNEGGFDFAFVDA
                     DKLNYWNYHERLIRLIKVGGIIVYDNTLWGGSVAEPDSSTPEWRIEVKKATLELNKKL
                     SADQRVQISQAALGDGITICRRLY"
     intron          28370. .28416
                     /gene="T25K17.30"
                     /number=1
```

```
exon            28417. .28641
                /gene="T25K17.30"
                /number=2
intron          28642. .28747
                /gene="T25K17.30"
                /number=2
exon            28748. .28879
                /gene="T25K17.30"
                /number=3
intron          28880. .28965
                /gene="T25K17.30"
                /number=3
exon            28966. .29256
                /gene="T25K17.30"
                /number=4
tRNA            29533. .29604
                /note="Pro-tRNA_(TGG)"
                /product="tRNA-Pro"
exon            30136. .30215
                /gene="T25K17.40"
                /number=1
gene            30136. .30663
                /gene="T25K17.40"
CDS             join(30136. .30215,30384. .30663)
                /gene="T25K17.40"
                /note="strong similarity to ribosomal protein L31,
                Nicotiana glutinosa, U23784
                contains EST gb:H77007"
                /codon_start=1
                /product="putative ribosomal protein"
                /protein_id="CAB38952.1"
                /db_xref="GI:4539419"
                /translation="MSEKKGRKEEVVTREYTINLHRRLHSCTFKKKAPKAIKEIRKFA
                EKEMGTKDVRVDVKLNKQIWSKGIRGPPRRIRVRVARKRNDDEDAKEEFFSLVTVAEI
                PAEGLSGLGTKIIEEED"
intron          30216. .30383
                /gene="T25K17.40"
                /number=1
exon            30384. .30663
                /gene="T25K17.40"
                /number=2
misc_feature    30910. .31193
                /note="EST AA586104 matches to position 30910-31193; no
                obvious genmodel"
gene            32558. .33006
                /gene="T25K17.50"
exon            complement(32558. .32659)
                /gene="T25K17.50"
                /number=1
gene            complement(join(32558. .32659,32745. .32806,32898. .33006))
                /gene="T25K17.50"
CDS             complement(join(<32558. .32659,32745. .32806,32898. .33006))
                /gene="T25K17.50"
                /note="contains EST gb:F45349"
                /codon_start=1
                /product="putative protein (fragment)"
                /protein_id="CAB38953.1"
                /db_xref="GI:4539420"
                /translation="MKKLARKWRRTRRDEDDDKFVLPTSEDVDSRPIDTQEQEEYVRY
                LEEAHAQQSLQWRSVFVVLLICFGAFLFYSIFQQFMSPWELVLIISP"
intron          complement(32660. .32744)
                /gene="T25K17.50"
                /number=1
exon            complement(32745. .32806)
                /gene="T25K17.50"
                /number=2
intron          complement(32807. .32897)
                /gene="T25K17.50"
                /number=2
exon            complement(32898. .33006)
                /gene="T25K17.50"
                /number=3
gene            33751. .35163
                /gene="T25K17.60"
CDS             join(33751. .34158,34325. .34648,34729. .34863,34945. .35163)
                /gene="T25K17.60"
                /note="strong similarity to WSI76 protein induced by water
                stress, Oryzasativa, D26537"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB38954.1"
                /db_xref="GI:4539421"
                /translation="MAQMSMTVEKSIKADVTVSHDRVKRAYVTFLAGNKDYWMGVVGL
                AKGLRKVKSAYPLVVAILPDVPEEHRQILLAQGCIIREIEPVYPPENKTGYSMAYYVI
```

Query Match 4.1%; Score 54; DB 8; Length 89904;
Best Local Similarity 56.6%; Pred. No. 0.088;
Matches 120; Conservative 0; Mismatches 90; Indels 2; Gaps 1;

```
Qy    259 AGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATT 318
          | || |  |  ||| |   |     | || ||||  ||| | |   |  |||  ||  ||
Db  25248 ATTAATATTTAATATAATAATTACTAAATAATCATTTATTTGAATTTTGTTTATTATTTT 25307

Qy    319 AAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAA 378
          | |||||| ||||  |   || |  | ||| |  |   | ||| |||  |||||| |||
Db  25308 CATTTAAAATTAAAGTTCTGTATAAATTTAAAATAATTTATATTATAAATTTTTCTTCAA 25367

Qy    379 CAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTA 438
          |  ||||  |||| || |  ||   |||| |||| | |  | |   || |||   |
Db  25368 AACATTTATTTATTTTATA--GTAAAATTTTCAATTATATTACAAAGTTAAAATAAGTTT 25425

Qy    439 CAATGGAAAACAGTGCTTTGTACGACTATATA 470
          || |  |||||| |  ||| ||| | | || |
Db  25426 CACTTTAAAACATTTATTTATACTAATTTAAA 25457
```

```
RESULT 12
ATCHRIV64
LOCUS       ATCHRIV64              196286 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 64.
ACCESSION   AL161564
VERSION     AL161564.2  GI:7269427
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 50686 to 155423; 144355 to 196286)
  AUTHORS   Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 77316 to 77804)
  AUTHORS   Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 82307)
  AUTHORS   Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 196286)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV63 at the 5' end and an
            overlap with ATCHRIV65 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .196286
```

```
                /organism="Arabidopsis thaliana"
                /variety="Columbia"
                /db_xref="taxon:3702"
                /chromosome="4"
     gene       6687. .8444
                /gene="AT4g25770"
     exon       complement(6687. .6833)
                /gene="AT4g25770"
                /number=1
     gene       complement(join(6687. .6833,6920. .7036,7083. .7163,
                7278. .7370,7494. .7733,7782. .7882,7945. .8042,8134. .8444))
                /gene="AT4g25770"
     CDS        complement(join(6687. .6833,6920. .7036,7083. .7163,
                7278. .7370,7494. .7733,7782. .7882,7945. .8042,8134. .8444))
                /gene="AT4g25770"
                /note="similarity to hypothetical protein T27I1.6 -
                Arabidopsis thaliana, PIR2:T00623"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB79432.1"
                /db_xref="GI:7269428"
                /translation="MIASFSRFQISIPDLIPSSPRFLPPSQNQFRFRLYLQFPAQPKN
                LCQMEWMNRLGIGCFAANRSKRELTTEIDNGGEDFFDADVMESAEKPDHLVVMVNGIV
                GSAADWKYAAEQFVKKFPDKVLVHRACFLTYRYCGSESNSATLTFDGVDKMGERLANE
                VVFGCIVVSSVLGVVKHRSGLKKISFVAHSLGGLVARYAIGKLYEQPGEVDSLDSPSK
                EKSARGGEIAGLEPMNFITFATPHLGSRGHRQFPILCGLPFLERTASQTAHLAAGRTG
                KHLFLIDMLIISSYDLNRSALNAFKRRVAYANRLSMVFDYFYHLQEFALTNEKAMVGW
                RTSSIRRPNELPKPNLLATDPNYPHIVYVERGNVDNGSCQSTSTVVTEQDTDLEGFAL
                FLT"
     intron     complement(6834. .6919)
                /gene="AT4g25770"
                /number=1
     exon       complement(6920. .7036)
                /gene="AT4g25770"
                /number=2
     intron     complement(7037. .7082)
                /gene="AT4g25770"
                /number=2
     exon       complement(7083. .7163)
                /gene="AT4g25770"
                /number=3
     intron     complement(7164. .7277)
                /gene="AT4g25770"
                /number=3
     exon       complement(7278. .7370)
                /gene="AT4g25770"
                /number=4
     intron     complement(7371. .7493)
                /gene="AT4g25770"
                /number=4
     exon       complement(7494. .7733)
                /gene="AT4g25770"
                /number=5
     intron     complement(7734. .7781)
                /gene="AT4g25770"
                /number=5
     exon       complement(7782. .7882)
                /gene="AT4g25770"
                /number=6
     intron     complement(7883. .7944)
                /gene="AT4g25770"
                /number=6
     exon       complement(7945. .8042)
                /gene="AT4g25770"
                /number=7
     intron     complement(8043. .8133)
                /gene="AT4g25770"
                /number=7
     exon       complement(8134. .8444)
                /gene="AT4g25770"
                /number=8
     CDS        9569. .10141
                /gene="AT4g25780"
                /note="strong similarity to gene PR-1 protein - Medicago
                truncatula, PIR2:S47171
                Contains Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7
                signatures AA147-157;Extracellular proteins
                SCP/Tpx-1/Ag5/PR-1/Sc7 signatures AA173-184
                contains EST gb:z26548, AA586201, AA712838, AA712715"
                /codon_start=1
                /product="putative pathogenesis-related protein"
                /protein_id="CAB79433.1"
                /db_xref="GI:7269429"
                /translation="MSSSYLRVILLLGALNVVVSLSITNSLITKSATLGQVFRICKNL
                CPGCDHDSLQFLFRHNLVRAARFEPPLIWDRRLQNYAQGWANQRRGDCALRHSVSNGE
                FNLGENIYWGYGANWSPADAVVAWASEKRFYHYGSNTCDAGQMCGHYTQIVWKSTRRV
                GCARVVCDNGGIFMTCNYDPPGNYIGQKPY"
     exon       9569. .10141
                /gene="AT4g25780"
                /number=1
     gene       9569. .10141
                /gene="AT4g25780"
     exon       complement(10710. .11074)
                /gene="AT4g25790"
                /number=1
     gene       10710. .11752
                /gene="AT4g25790"
     gene       complement(join(10710. .11074,11485. .11752))
                /gene="AT4g25790"
     CDS        complement(join(10710. .11074,11485. .11752))
                /gene="AT4g25790"
                /note="strong similarity to gene PR-1 protein - Medicago
                truncatula, PIR2:S47171
                Contains Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7
                signatures AA166-176;Extracellular proteins
                SCP/Tpx-1/Ag5/PR-1/Sc7 signatures AA193-204
                contains EST gb:AA712715, N65358"
                /codon_start=1
                /product="putative pathogenesis-related protein"
                /protein_id="CAB79434.1"
                /db_xref="GI:7269430"
                /translation="MATTTSTQAIVFAIALFLVAIQAVHADYYRHRPQRPITPSPPPY
                VAKPQPLPTPSPKPILYQPPPTYQPPTGSFEQQFLDPHNTVRGGLGLPPLVWDVKIAS
                YATWWANQRRYDCSLTHSTGPYGENLFWGSGSDFTSTFAVESWTVEAKSYNHMTNTCE
                GDGMCGHYTQIVWRETRRLGCARVVCENGAGVFITCNYDPPGNYVGEKPY"
     intron     complement(11075. .11484)
                /gene="AT4g25790"
                /number=1
     exon       complement(11485. .11752)
                /gene="AT4g25790"
                /number=2
     gene       13985. .16227
                /gene="AT4g25800"
     CDS        join(13985. .14010,14239. .14669,14761. .14921,15016. .15219,
                15390. .15595,15645. .16227)
                /gene="AT4g25800"
                /note="EST GB:F14385 in 5' untranslated region
                strong similarity to calmodulin-binding protein (TCB60) -
                Nicotianatabacum, PID:g1698548
                contains EST gb:F14386, F14385, N96168, N96141"
                /codon_start=1
                /product="putative calmodulin-binding protein"
                /protein_id="CAB79435.1"
                /db_xref="GI:7269431"
                /translation="MNQFCLLVRLSLPLFTGGRVEGEQGATIHVVLIDANTGRPVTVG
                PEASLKLEVVVLGGDFNNEDDEDWTQEEFESHVVKEREGKRPLLTGDLFVVLKEGVGT
                LGEIVFTDNSSWIRSRKFRLGLRVPSGYCDGIRIREAKTEAFSVKDHRGELYKKHYPP
                ALNDEVWRLEKIGKDGAFHKRLTAAGIVTVEGFLRQLVRDSTKLRAILGSGMSNKMWD
                LLVEHAKTCVLSGKLYIYYTEDSRSVGVVFNNIYELSGLITEDQYLSADSLSESQKVY
                VDGLVKKAYENWNQVVEYEGESLLNLNQPERLDISQTDPVTALASYSTVPLSQFPEFA
                IEGWQISPFLYSTSIFTAGYNQTLTTALPHNPQAQFDFVPQQDQFIGIQQPQTQTNIE
                NENVTRLVLGPPQSSTGGYQDIKSSADQENLNPFEDWTNLSENDFFSEEEIRQTSHDL
                LANEDMQQLLFSMGGGKGEDGFTFPSFMQNTPMMQGYDEEGRGRSGKAVVGWLKVKAA
                MRWGFFIRRKAAERRAQIVELHDNNDGK"
     exon       13985. .14010
                /gene="AT4g25800"
```

```
                     /number=1
     intron          14011. .14238
                     /gene="AT4g25800"
                     /number=1
     exon            14239. .14669
                     /gene="AT4g25800"
                     /number=2
     intron          14670. .14760
                     /gene="AT4g25800"
                     /number=2
     exon            14761. .14921
                     /gene="AT4g25800"
                     /number=3
     intron          14922. .15015
                     /gene="AT4g25800"
                     /number=3

  Query Match             4.1%;  Score 54;  DB 8;  Length 196286;
  Best Local Similarity   56.6%;  Pred. No. 0.11;
  Matches  120;  Conservative    0;  Mismatches   90;  Indels    2;  Gaps    1;

Qy     259 AGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATT 318
           | || |  |  ||| |   |     | || ||||  ||| | |   |  |||  ||  ||
Db  169602 ATTAATATTTAATATAATAATTACTAAATAATCATTTATTTGAATTTTGTTTATTATTTT 169661

Qy     319 AAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAA 378
           | |||||| ||||  |   ||  |  | ||| |  |   | ||| |||  |||||| |||
Db  169662 CATTTAAAATTAAAGTTCTGTATAAATTTAAAATAATTTATATTATAAATTTTTCTTCAA 169721

Qy     379 CAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTA 438
           |  ||||  |||| || |  ||   ||||  |||| | |  | |   || |||   |
Db  169722 AACATTTATTTATTTTATA--GTAAAATTTTCAATTATATTACAAAGTTAAAATAAGTTT 169779

Qy     439 CAATGGAAAACAGTGCTTTGTACGACTATATA 470
           || |  ||||||  |  ||| ||| | | || |
Db  169780 CACTTTAAAACATTTATTTATACTAATTTAAA 169811


RESULT 13
AC006248/c
LOCUS       AC006248              115522 bp    DNA     linear   PLN 05-APR-2000
DEFINITION  Arabidopsis thaliana chromosome II section 91 of 255 of the
            complete sequence. Sequence from clones F26H6, F9O13, F19G14.
ACCESSION   AC006248 AE002093
VERSION     AC006248.3  GI:6598542
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 115522)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20083487
   PUBMED   10617197
REFERENCE   2  (bases 1 to 115522)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:4335711.
            The sequence and annotation of chromosome 2 were merged from those
            of the individual clones on this chromosome after removing
            overlaps. For detailed information, please see the TIGR web site
            (http://www.tigr.org/tdb/at/at.html).

            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL
            (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
            University of Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/services/NetGene2/), searches of the
            complete sequence against a peptide database and plant EST
            databases at TIGR, and manual curations based on those analyses.
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by two
            or more gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
            numbered from the top to bottom of the chromosome.

            We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
            F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
            F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
            and Satoshi Tabata for helpful assistance. In addition, we would
            like to thank the TIGR Bioinformatics Department, especially Lixin
            Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
            Peterson, Michael Holmes, and Delwood Richardson for software and
            database support.

            This work was supported by the National Science Foundation,
            Department of Energy and the US Department of Agriculture.

            Address all correspondence to:  at@tigr.org.
FEATURES            Location/Qualifiers
     source         1. .115522
                    /organism="Arabidopsis thaliana"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="II"
     misc_feature   complement(<1. .1408)
                    /note="Sequence from clone F26H6"
     repeat_region  515. .628
                    /rpt_family="(TAAAA)n"
     mRNA           join(<940. .1083,1286. .1597,1804. .2328,2426. .>2653)
                    /gene="At2g15460"
     gene           <940. .>2653
                    /gene="At2g15460"
                    /note="F26H6.2"
     CDS            join(940. .1083,1286. .1597,1804. .2328,2426. .2653)
                    /gene="At2g15460"
                    /codon_start=1
                    /product="putative polygalacturonase"
                    /protein_id="AAD17390.2"
                    /db_xref="GI:6598543"
                    /translation="MACISFAFKALCFCLFFIVVASRSTVRPKVFNVQRHGAKPDGKT
                    DNANAFTSIWSRACKRISGSSKIYVPKGTFYLGGVEFVGPCKNPIEFIIDGTLLAPAN
                    PSDIKQDTWINFRYINNLSISGSGTLDGQGKQSWPHNDCHTNPNCPKLAMTMGFAFVN
                    NSNIKDITSLNSKMGHFNFFSVHHFNITGVTITAPGDSPNTDGIKMGSCSNIQISDTN
                    IGTGDDCIAILSGTTNLNISNVNCGPGHGISVGSLGKSKDEKDVKDLIVRDVIFNGTS
                    DGIRIKTWESSASKILVSNFVYENIQMIDVGKPINIDQKYCPHPPCEHERKGESHVQI
                    QNLKLKNIYGTSKNKVAVNLQCSKIFPCKNVELIDINIKQNGVKDGSSTSVCENVDGF
                    ARGKMFPRIV"
     misc_feature   1409. .115321
                    /note="Sequence from clone F9O13"
     repeat_region  complement(2357. .2378)
                    /rpt_family="(TAAAA)n"
     repeat_region  3911. .4024
                    /rpt_family="(TAAAA)n"
     mRNA           join(<4336. .4479,4682. .4993,5202. .5726,5830. .>6057)
```

```
                /gene="At2g15470"
gene            <4336. .>6057
                /gene="At2g15470"
                /note="F9O13.2"
CDS             join(4336. .4479,4682. .4993,5202. .5726,5830. .6057)
                /gene="At2g15470"
                /codon_start=1
                /product="putative polygalacturonase"
                /protein_id="AAD17391.1"
                /db_xref="GI:4335713"
                /translation="MACISFAFKALCFCLFFIVVASRSTVRPKVFNVQRHGAKPDGKT
                DNANAFTSIWSRACKRISGSSKIYVPKGTFYLGGVEFVGLCKNPIEFIIDGTLLAPAN
                PSDIKQDTWINFRYINNLSISGSGTLDGQGKQSWPHNDCHTNPNCPKLAMTMGFAFVN
                NSNIKDITSLNSKMGHFNFFSVHHFNITGVTITAPGDSPNTDGIKMGSCSNIQISDTN
                IGTGDDCIAILSGTTNLNISNVNCGPGHGISVGSLGKNKDEKDAKDLIVRDVIFNGTS
                DGIRIKTWESSASKILVSNFVYENIQMIDVGKPINIDQKYCPHPPCEHERKSHVQIQN
                LKLKNIYGTSKNKVAMNLQCSKIFPCKNVELIDINIKQNGVKDGSSTSVCENVDGFAR
                GKMFPQHCLN"
repeat_region   complement(5755. .5776)
                /rpt_family="(TAAAA)n"
repeat_region   7220. .7281
                /rpt_family="(TA)n"
repeat_region   complement(7698. .7802)
                /rpt_family="(TA)n"
repeat_region   complement(7735. .7838)
                /rpt_family="(TAAA)n"
repeat_region   complement(8026. .8142)
                /rpt_family="(TAA)n"
repeat_region   complement(8080. .8183)
                /rpt_family="(GAAAA)n"
repeat_region   complement(8739. .8780)
                /rpt_family="(CAAAA)n"
mRNA            join(<8823. .9813,9995. .>10458)
                /gene="At2g15480"
gene            <8823. .>10458
                /gene="At2g15480"
                /note="F9O13.3"
CDS             join(8823. .9813,9995. .10458)
                /gene="At2g15480"
                /codon_start=1
                /product="putative glucosyltransferase"
                /protein_id="AAD17392.1"
                /db_xref="GI:4335714"
                /translation="MNREVSERIHILFFPFMAQGHMIPILDMAKLFSRRGAKSTLLTT
                PINAKIFEKPIEAFKNQNPDLEIGIKIFNFPCVELGLPEGCENADFINSYQKSDSGDL
                FLKFLFSTKYMKQQLESFIETTKPSALVADMFFPWATESAEKLGVPRLVFHGTSFFSL
                CCSYNMRIHKPHKKVATSSTPFVIPGLPGDIVITEDQANVAKEETPMGKFMKEVRESE
                TNSFGVLVNSFYELESAYADFYRSFVAKRAWHIGPLSLSNRELGEKARRGKKANIDEQ
                ECLKWLDSKTPGSVVYLSFGSGTNFTNDQLLEIAFGLEGSGQSFIWVVRKNENQGDNE
                EWLPEGFKERTTGKGLIIPGWAPQVLILDHKAIGGFVTHCGWNSAIEGIAAGLPMVTW
                PMGAEQFYNEKLLTKVLRIGVNVGATELVKKGKLISRAQVEKAVREVIGGEKAEERRL
                WAKKLGEMAKAAVEEGGSSYNDVNKFMEELNGRK"
mRNA            join(<11828. .12737,12932. .>13404)
                /gene="At2g15490"
gene            <11828. .>13404
                /gene="At2g15490"
                /note="F9O13.4"
CDS             join(11828. .12737,12932. .13404)
                /gene="At2g15490"
                /codon_start=1
                /product="putative glucosyltransferase"
                /protein_id="AAD17393.1"
                /db_xref="GI:4335715"
                /translation="MAKLFARRGAKSTLLTTPINAKILEKPIEAFKVQNPDLEIGIKI
                LNFPCVELGLPEGCENRDFINSYQKSDSFDLFLKFLFSTKYMKQQLESFIETTKPSAL
                VADMFFPWATESAEKIGVPRLVFHGTSSFALCCSYNMRIHKPHKKVASSSTPFVIPGL
                PGDIVITEDQANVTNEETPFGKFWKEVRESETSSFGVLVNSFYELESSYADFYRSFVA
                KKAWHIGPLSLSNRGIAEKAGRGKKANIDEQECLKWLDSKTPGSVVYLSFGSGTGLPN
                EQLLEIAFGLEGSGQNFIWVVSKNENQVGTGENEDWLPKGFEERNKGKGLIIRGWAPQ
                VLILDHKAIGGFVTHCGWNSTLEGIAAGLPMVTWPMGAEQFYNEKLLTKVLRIGVNVG
                ATELVKKGKLISRAQVEKAVREVIGGEKAEERRLRAKELGEMAKAAVEEGGSSYNDVN
                KFMEELNGRK"
mRNA            join(<13713. .13796,13945. .14093,14268. .14303,14375. .14413,
                14667. .>14757)
                /gene="At2g15500"
gene            <13713. .>14757
                /gene="At2g15500"
                /note="F9O13.5"
CDS             join(13713. .13796,13945. .14093,14268. .14303,14375. .14413,
                14667. .14757)
                /gene="At2g15500"
                /codon_start=1
                /product="putative poly(A)-binding protein"
                /protein_id="AAD17394.1"
                /db_xref="GI:4335716"
                /translation="MSKLIGMLLNDKQVYVGPFLRGKKGILLAVESLKGNTFDDKQWV
                VDRAQKKSEREMELKVPYEQSLKEIADKFQSYNFRGSGFVAFSTQNVTISGICKDVID
                ESVAMPPSVGPRMPIYPLGSPGIRKSMFYG"
mRNA            complement(join(<16628. .18427,18538. .18661,18914. .19100,
                19343. .19660,19695. .>20682))
                /gene="At2g15510"
gene            complement(<16628. .>20682)
                /gene="At2g15510"
                /note="F9O13.6"
```

```
  Query Match             4.1%;  Score 53.8;  DB 8;  Length 115522;
  Best Local Similarity   53.3%;  Pred. No. 0.1;
  Matches  136;  Conservative    0;  Mismatches  117;  Indels    2;  Gaps    1;

Qy     226 TGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACAT 285
           ||  |||| | || ||  |  |      |  |  | ||||  |  | | |   |     |
Db    8079 TGATTATTGTAATAGTAATCAAAAAGGTGAATTAATAGTATTTAACATAATAATTAGTAA 8020

Qy     286 ATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTT 345
           |  ||||  ||| | |    |  |||  ||| || | |||||| |||| |    | | |
Db    8019 AAAATCATTTATTTGAATTTTGTTTATTAATTTCATTTAAAATTAAAGTTCTGCATAAAT 7960

Qy     346 CTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTA 405
            ||| |  ||  | ||| |||  || ||| ||| |  |||| |||| || | |     |
Db    7959 TTAAAATAAATTATATTATAAATTTCTCTTCAAAACATTTATTTATTTTATAATAAAAAA 7900

Qy     406 TTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGTACGACT 465
           ||   |||| | |  | |   || |||   | || |  |||||| |  ||| ||| | |
Db    7899 TT--AAATTATATTACAAATTTAAAATAAGTTTCACTTTAAAACATTTATTTATACTAAT 7842

Qy     466 ATATACGCGATGTGT 480
           || |    || | |
Db    7841 TTAAAATAAATTTAT 7827
```

```
RESULT 14
AC105576/c
LOCUS       AC105576              214033 bp    DNA     linear   HTG 09-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-236F15, WORKING DRAFT SEQUENCE, 50
            unordered pieces.
ACCESSION   AC105576
VERSION     AC105576.1  GI:18092798
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 214033)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Bryant,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
```

```
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 214033)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GNFN
                Center clone name: CH230-236F15
            -------------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 185044 bases at least Q40
                Consensus quality: 190296 bases at least Q30
                Consensus quality: 194221 bases at least Q20
                Estimated insert size: 190368; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 50 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    16332: contig of 16332 bp in length
            *    16333    16432: gap of unknown length
            *    16433    26891: contig of 10459 bp in length
            *    26892    26991: gap of unknown length
            *    26992    39114: contig of 12123 bp in length
            *    39115    39214: gap of unknown length
            *    39215    47706: contig of 8492 bp in length
            *    47707    47806: gap of unknown length
            *    47807    54905: contig of 7099 bp in length
            *    54906    55005: gap of unknown length
            *    55006    61730: contig of 6725 bp in length
            *    61731    61830: gap of unknown length
            *    61831    68922: contig of 7092 bp in length
            *    68923    69022: gap of unknown length
            *    69023    73955: contig of 4933 bp in length
            *    73956    74055: gap of unknown length
            *    74056    79739: contig of 5684 bp in length
            *    79740    79839: gap of unknown length
            *    79840    87203: contig of 7364 bp in length
            *    87204    87303: gap of unknown length
            *    87304    95716: contig of 8413 bp in length
            *    95717    95816: gap of unknown length
            *    95817   101220: contig of 5404 bp in length
            *   101221   101320: gap of unknown length
            *   101321   106991: contig of 5671 bp in length
            *   106992   107091: gap of unknown length
            *   107092   111184: contig of 4093 bp in length
            *   111185   111284: gap of unknown length
            *   111285   117002: contig of 5718 bp in length
            *   117003   117102: gap of unknown length
            *   117103   122719: contig of 5617 bp in length
            *   122720   122819: gap of unknown length
            *   122820   127676: contig of 4857 bp in length
            *   127677   127776: gap of unknown length
            *   127777   132248: contig of 4472 bp in length
            *   132249   132348: gap of unknown length
            *   132349   137419: contig of 5071 bp in length
            *   137420   137519: gap of unknown length
            *   137520   140519: contig of 3000 bp in length
            *   140520   140619: gap of unknown length
            *   140620   145613: contig of 4994 bp in length
            *   145614   145713: gap of unknown length
            *   145714   150743: contig of 5030 bp in length
            *   150744   150843: gap of unknown length
            *   150844   153229: contig of 2386 bp in length
            *   153230   153329: gap of unknown length
            *   153330   155705: contig of 2376 bp in length
            *   155706   155805: gap of unknown length
            *   155806   159641: contig of 3836 bp in length
            *   159642   159741: gap of unknown length
            *   159742   161937: contig of 2196 bp in length
            *   161938   162037: gap of unknown length
            *   162038   166136: contig of 4099 bp in length
            *   166137   166236: gap of unknown length
            *   166237   169654: contig of 3418 bp in length
            *   169655   169754: gap of unknown length
            *   169755   173209: contig of 3455 bp in length
            *   173210   173309: gap of unknown length
            *   173310   174872: contig of 1563 bp in length
            *   174873   174972: gap of unknown length
            *   174973   178370: contig of 3398 bp in length
            *   178371   178470: gap of unknown length
            *   178471   181766: contig of 3296 bp in length
            *   181767   181866: gap of unknown length
            *   181867   184177: contig of 2311 bp in length
            *   184178   184277: gap of unknown length
            *   184278   186253: contig of 1976 bp in length
            *   186254   186353: gap of unknown length
            *   186354   188889: contig of 2536 bp in length
            *   188890   188989: gap of unknown length
            *   188990   190730: contig of 1741 bp in length
            *   190731   190830: gap of unknown length
            *   190831   192237: contig of 1407 bp in length
            *   192238   192337: gap of unknown length
            *   192338   194466: contig of 2129 bp in length
            *   194467   194566: gap of unknown length
            *   194567   196224: contig of 1658 bp in length
            *   196225   196324: gap of unknown length
            *   196325   197794: contig of 1470 bp in length
            *   197795   197894: gap of unknown length
            *   197895   199371: contig of 1477 bp in length
            *   199372   199471: gap of unknown length
            *   199472   202264: contig of 2793 bp in length
```

```
            *   202265   202364: gap of unknown length
            *   202365   204373: contig of 2009 bp in length
            *   204374   204473: gap of unknown length
            *   204474   205726: contig of 1253 bp in length
            *   205727   205826: gap of unknown length
            *   205827   207670: contig of 1844 bp in length
            *   207671   207770: gap of unknown length
            *   207771   208820: contig of 1050 bp in length
            *   208821   208920: gap of unknown length
            *   208921   210151: contig of 1231 bp in length
            *   210152   210251: gap of unknown length
            *   210252   211652: contig of 1401 bp in length
            *   211653   211752: gap of unknown length
            *   211753   212869: contig of 1117 bp in length
            *   212870   212969: gap of unknown length
            *   212970   214033: contig of 1064 bp in length.
FEATURES             Location/Qualifiers
     source          1. .214033
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-236F15"
BASE COUNT    63123 a  41277 c  42183 g  62289 t   5161 others
ORIGIN

  Query Match             4.1%;  Score 53.6;  DB 2;  Length 214033;
  Best Local Similarity   50.8%;  Pred. No. 0.13;
  Matches  128;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy     815 CCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCC 874
           |||  |  | | |  | || ||    | || | | ||| | | | | | ||| ||   ||
Db  140327 CCCCCGAACCCCCGTCCCGCCCCGCCCCCCCTTCGTCTCCTGACCCCCCCCACCCCCCCC 140268

Qy     875 GTGTCGCGGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCC 934
           | |    |  | |  ||  |||   ||| | | |     |||||||  |||   ||
Db  140267 TCCCCTCCCCCCCCCCCCTCCCGCCCGCGCCCCCCCGCCCTCGCCCGTCCTCGCCCCCCC 140208

Qy     935 CGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGA 994
           |  |   |  | ||| |  |  || | |  ||  ||  |||| | || |    | |||
Db  140207 CCCGCTCGCTCCCGCCCCCCCCCGCGCCCTCCGTCGGGGCTGCGGCCGCCCCCCGCCCCG 140148

Qy     995 CACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCACGGCCGACTGC 1054
           ||||  || ||| | |    |||  |||    || |     || || ||  |  || |
Db  140147 TCCCCTCCCGCCCCCTCGCGCCTTCCGCCCCCCCCTCCCGCTCCCCTCCCCGCTCCCCTCC 140088

Qy    1055 GCTGCGCCGCCC 1066
           | ||||| |||
Db  140087 CCCGCGCCCCCC 140076


RESULT 15
AF254090
LOCUS       AF254090                 614 bp    DNA     linear   INV 06-DEC-2000
DEFINITION  Rhopalosiphum maidis cytochrome oxidase subunit II (COII) gene,
            partial cds; mitochondrial gene for mitochondrial product.
ACCESSION   AF254090
VERSION     AF254090.1  GI:11559553
KEYWORDS    .
SOURCE      corn leaf aphid.
  ORGANISM  Mitochondrion Rhopalosiphum maidis
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha;
            Aphidiformes; Aphidoidea; Aphididae; Aphidini; Rhopalosiphum.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Chen,Y., Giles,K.L., Payton,M.E. and Greenstone,M.H.
  TITLE     Identifying key cereal aphid predators by molecular gut analysis
  JOURNAL   Mol. Ecol. 9 (11), 1887-1898 (2000)
  MEDLINE   20546168
   PUBMED   11091324
REFERENCE   2  (bases 1 to 614)
  AUTHORS   Chen,Y., Giles,K.L., Payton,M.E. and Greenstone,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2000) Entomology and Plant Pathology, Oklahoma
            State University, 127/110 Noble Research Center, Stillwater, OK
            74078-3033, USA
FEATURES             Location/Qualifiers
     source          1. .614
                     /organism="Rhopalosiphum maidis"
                     /organelle="mitochondrion"
                     /db_xref="taxon:43146"
     gene            <1. .>614
                     /gene="COII"
     CDS             <1. .>614
                     /gene="COII"
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome oxidase subunit II"
                     /protein_id="AAG38002.1"
                     /db_xref="GI:11559554"
                     /translation="LIFFHDHTIFIIIMIMTTITYMMIFIMKNKFINIKISENQMIEF
                     IWTTTPPIILIFIAMPSLHLLYLMDEIKCPILTIKIFGHQWFWSYEYSDFSNIEFESY
                     MMNELNKENFRLIEVDNKTIIPFKFNIRLLISSDDVIHSWTIPSLAIKIDAIPGRMNQ
                     INLFMNRPGMYFGQCSEICGINHSFMPIQIESINLNKFIYWIKN"
BASE COUNT      247 a     69 c     48 g    250 t
ORIGIN

  Query Match             4.1%;  Score 53.4;  DB 3;  Length 614;
  Best Local Similarity   56.3%;  Pred. No. 0.038;
  Matches  120;  Conservative    0;  Mismatches   91;  Indels    2;  Gaps    1;

Qy     235 TTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACA 294
           ||| | || |||  |||| |||| | |   |  ||     ||    |||   | || ||
Db       1 TTAATTTTTTTTCATGATCATACAATTTTTATTATTATTATAATTATAACAACAATTACT 60

Qy     295 TATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCA 354
           ||||| | |  |||||| | |   ||| ||||  ||  |||    |||||||| | |
Db      61 TATATAATA--ATTTTTATTATAAAAAATAAATTTATTAATATTAAAATTTCTGAAAATC 118

Qy     355 ACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATT 414
           |    |||| | |  |||   | ||||||  |||||||| | | |  || |||| | |
Db     119 AAATAATTGAATTTATTTGAACTACAACTCCACCTATTATTCTTATTTTTATTGCTATAC 178

Qy     415 TCACTCTATAAACAACATAGTCTACAATGGAAA 447
           || || |   |  ||| |  | |   ||||
Db     179 CTTCTTTACATCTATTATACTTAATAGATGAAA 211


Search completed: November  7, 2002, 12:14:29
Job time : 3553.14 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:29:53 ; Search time 166.564 Seconds
(without alignments)
13503.300 Million cell updates/sec

Title: US-09-905-558C-3
Perfect score: 1310
Sequence: 1 cccatcgctgctttgtctac.....caagggaaagtgatcggatg 1310

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	235	17.9	12313	21	AAZ35184	Corn delta-12 desa
	2	232.8	17.8	2126	24	AAS96571	Corn promoter sequ
c	3	124.6	9.5	1888	22	AAF81468	Corn promoter clon
	4	50.2	3.8	6327	22	AAS45444	Chemically pretrea
c	5	50	3.8	10133	24	ABL32458	Human immune syste
c	6	49.4	3.8	73334	24	ABL34124	Human immune syste
c	7	49.2	3.8	11691	24	ABL34241	Human immune syste
	8	48.8	3.7	1337	20	AAZ17263	Human gene express
c	9	48	3.7	399	22	AAH93294	Plasmodium falcipa
c	10	47.8	3.6	326	22	AAS56505	Human cDNA for an
c	11	47.8	3.6	73334	24	ABL34125	Human immune syste
c	12	47.4	3.6	6627	24	ABL33574	Human immune syste
c	13	47.2	3.6	6033	21	AAA70152	Plasmodium falcipa
	14	46.8	3.6	281	22	AAL16482	Human breast cance
c	15	46.8	3.6	6107	24	AAS61315	Human gene regulat
	16	46.6	3.6	20420	22	AAK73165	Human immune/haema
c	17	46.4	3.5	463	22	AAS56574	Human cDNA for an
c	18	46	3.5	6988	24	ABL34440	Human immune syste
c	19	45.8	3.5	2140	21	AAZ49411	Pea DRR206 protein
c	20	45.8	3.5	8991	24	AAS61295	Human gene regulat
	21	45.6	3.5	3730	22	ABA19745	Human nervous syst
c	22	45.6	3.5	6699	24	AAS63321	Chemically pretrea
c	23	45.6	3.5	7049	24	ABL32157	Human immune syste
	24	45.2	3.5	4257	19	AAV68520	The nucleotide seq
	25	45.2	3.5	4257	19	AAV10362	Infected cell prot
c	26	45.2	3.5	12001	16	AAQ76213	HSV L/ST region.
	27	45.2	3.5	13326	24	ABL33712	Human immune syste
c	28	45.2	3.5	17918	24	AAS61418	Human gene regulat
c	29	45	3.4	687	24	ABL34283	Human immune syste
	30	45	3.4	883	22	AAL26436	Human breast cance
	31	45	3.4	883	22	AAL26587	Human breast cance
	32	45	3.4	883	22	AAL26703	Human breast cance
	33	45	3.4	883	22	AAL26709	Human breast cance
	34	45	3.4	883	22	AAL26724	Human breast cance
	35	45	3.4	883	22	AAL26756	Human breast cance
	36	45	3.4	883	22	AAL26774	Human breast cance
	37	45	3.4	883	22	AAL26785	Human breast cance
c	38	45	3.4	6478	22	AAS45417	Chemically pretrea
c	39	45	3.4	6626	22	AAS46810	Tumour suppressor
	40	45	3.4	10329	24	ABL34123	Human immune syste
c	41	45	3.4	16217	24	ABL32625	Human immune syste
	42	44.8	3.4	259	22	AAL25325	Human breast cance
c	43	44.8	3.4	11473	24	ABL33355	Human immune syste
c	44	44.8	3.4	17419	22	AAS45393	Chemically pretrea
c	45	44.8	3.4	17419	24	ABL33295	Human immune syste

ALIGNMENTS

RESULT 1
AAZ35184
ID AAZ35184 standard; DNA; 12313 BP.
XX
AC AAZ35184;
XX
DT 13-MAR-2000 (first entry)
XX
DE Corn delta-12 desaturase fad2-2 gene 5' untranslated region.
XX
KW Corn; maize; transgenic plant; lipid; food; feedstuff;
KW vegetable oil; seed oil; oleic acid; fatty acid desaturase;
KW delta-12 desaturase; fad2-2; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT intron 5651..12301
FT /*tag= a
FT TATA_signal 5439..5444
FT /*tag= b
XX
PN WO9964579-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-US12884.
XX
PR 11-JUN-1998; 98US-0088987.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX

```
PI   Shen JB;
XX
DR   WPI; 2000-097535/08.
XX
PT   New maize oleosin promoter, used for producing transgenic plants with
PT   altered fatty acid composition of the oil, used e.g. in animal feeds -
XX
PS   Example 2; Page 80-84; 108pp; English.
XX
CC   This is the nucleotide sequence of the 5' untranslated region of
CC   the novel corn fad2-2 gene that codes for delta-12 desaturase (see
CC   AAY32344). The invention generally relates to the preparation and use
CC   of nucleic acid fragments comprising all, or substantially all, of a
CC   corn oleosin promoter (see AAZ35165-77), a stearoyl-ACP desaturase
CC   (see AAZ35179-80) and a delta-12 desaturase, which can be used
CC   individually or in combination to modify the lipid profile of corn.
CC   Suppression of delta-12 desaturase expression can increase the
CC   oleic acid content of the seed oil. Also claimed are seeds of
CC   such plants, oil obtained from the grain of such plants, animal
CC   feed, use of the oil in food, feed, and cooking oil or industrial
CC   applications. The promoter region of fad2-2 may be used to
CC   express a gene of interest in transgenic corn plants.
XX
SQ   Sequence 12313 BP; 2845 A; 2969 C; 3170 G; 3329 T; 0 other;

  Query Match             17.9%;  Score 235;  DB 21;  Length 12313;
  Best Local Similarity   82.5%;  Pred. No. 3.3e-50;
  Matches  293;  Conservative    0;  Mismatches   60;  Indels    2;  Gaps    2;

Qy       5 TCGCTGCTTTGTCTACATCATGTTCTTCATCATCCTCCCCAGGCGACGCGTGCTGCTGTT 64
           ||||    | ||| ||| | || | ||| ||  |  ||| ||  ||||||||||||||||
Db    3444 TCGCGTGATCGTCCACAGCTTGCTGTTCGTCGCCTACCCAAGTTGACGCGTGCTGCTGTT 3503

Qy      65 CTTATTCA-GACTACCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGT 123
           ||| |||  | | |||||||||| |||||||  ||||||| ||| ||||| |||||| ||
Db    3504 CTTCTTCCCGGCGACCGTTCGAGGGACTGCACTGCGTACACCTTCCTGCACCGACTTCGT 3563

Qy     124 ACGGCTACATCGAACATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAA 183
           ||| ||||||||||||| | ||||||||||||||||||||||| ||| ||||  |||||| |
Db    3564 ACGACTACATCGAACAAACACACGAGATGTCTCGTGTGAATGGAGCCACTGGTGCCTTGA 3623

Qy     184 GCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTG 243
           ||||||||  |||||  |||||||  |||||||||| ||||||||||   ||| |||| ||
Db    3624 GCATCGGTCCCTCCGCTGGGTACACTCTGTTCTTCGTATTTGTGCAT-GTTTCATTGCTG 3682

Qy     244 TTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACA 303
           |||||||  ||||  |||||||||||||||  |||||||||||||||  |||||||||||||||
Db    3683 TTTACTGCTTATGCGAGTAGTTATACACACATGCACATACATGTCATCACATATATCACA 3742

Qy     304 ATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGA 358
             |  ||||||  |||||||||||||||||||||  ||||  ||||||||||   |||
Db    3743 CTGATTTTCTGGATTAAATTAAAACTAAAAATGCCTAACTTTCTAACACGTCCGA 3797


RESULT 2
AAS96571
ID   AAS96571 standard; DNA; 2126 BP.
XX
AC   AAS96571;
XX
DT   26-FEB-2002  (first entry)
XX
DE   Corn promoter sequence #14.
XX
KW   Corn; male reproductive tissue; plant regulatory sequence; Zea mays;
KW   promoter; transcription regulation; operably linked gene; monocot;
KW   dicot; wheat anther; plant fertility; insect tolerance;
KW   pathogen tolerance; herbicide tolerance; ds.
XX
OS   Zea mays.
XX
PN   WO200183790-A2.
XX
PD   08-NOV-2001.
XX
PF   30-APR-2001; 2001WO-US13739.
XX
PR   01-MAY-2000; 2000US-201255P.
XX
PA   (MONS ) MONSANTO TECHNOLOGY LLC.
XX
PI   Conner TW,  Dubois P,  Malven M,  Masucci JD;
XX
DR   WPI; 2002-055481/07.
XX
PT   Novel promoters isolated from corn for controlling gene expression in
PT   male reproductive tissues, such as anthers, tassels, and to regulate
PT   transcription of target genes including genes for insect or pathogen
PT   tolerance -
XX
PS   Claim 1; Page 115-116; 121pp; English.
XX
CC   The present invention relates to the isolation of plant regulatory
CC   sequences from the male reproductive tissues of corn (Zea mays). The
CC   promoter sequences, fragments, regions or cis elements of the sequences,
CC   are capable of regulating transcription of an operably linked DNA
CC   sequence. The promoter sequences confer enhanced expression of operably
CC   linked genes in monocot or dicot male reproductive tissues, such as
CC   anthers, especially wheat anthers and is useful for regulating
CC   transcription of a DNA sequence, by operably linking the DNA sequence
CC   to the promoter. The promoter sequences are useful in plants to regulate
CC   transcription of target genes including genes for control of fertility,
CC   insect or pathogen tolerance and herbicide tolerance. They are also
CC   useful as probes or primers in nucleic acid hybridisation experiments.
CC   The promoter sequences can be used in hybridisation assays of other
CC   plant tissues to identify closely related or homologous genes and
CC   associated regulatory sequences. AAS96558-AAS96577 represent the
CC   corn promoter sequences of the present invention.
XX
SQ   Sequence 2126 BP; 607 A; 419 C; 433 G; 667 T; 0 other;

  Query Match             17.8%;  Score 232.8;  DB 24;  Length 2126;
  Best Local Similarity   83.0%;  Pred. No. 5.6e-50;
  Matches  289;  Conservative    0;  Mismatches   57;  Indels    2;  Gaps    2;

Qy       5 TCGCTGCTTTGTCTACATCATGTTCTTCATCATCCTCCCCAGGCGACGCGTGCTGCTGTT 64
           ||||    | ||| ||| | || | ||| ||  | |||| ||  ||||||||||||||||
Db      82 TCGCGTGATCGTCCACAGCTTGCTGTTCGTCGCCTTCCCAAGTTGACGCGTGCTGCTGTT 141

Qy      65 CTTATTCA-GACTACCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGT 123
           ||| |||  | | |||||||||| |||||||  |||||||||| ||||| |||||| ||
Db     142 CTTCTTCCCGGCGACCGTTCGAGGGACTGCACTGCGTACATCTTCCTGCACCGACTTCGT 201

Qy     124 ACGGCTACATCGAACATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAA 183
           ||||||||||||||| | ||||||||||||||||||||| ||| ||||  |||||| |
Db     202 ACGGCTACATCGAACAAACACACGAGATGTCTCGTGTGAATGGAGCCACTGGTGCCTTGA 261

Qy     184 GCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTG 243
           ||||||||  |||||  |||||||  |||||||||| ||||||||||  ||| |||| ||
Db     262 GCATCGGTCCCTCCGCTGGGTACACTCTGTTCTTCGTATTTGTGCAT-GTTTCATTGCTG 320

Qy     244 TTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACA 303
           |||||||  |||| |||||||||||||  | ||||||||||| ||||||||||||| ||
Db     321 TTTACTGCTTATGCGAGTAGTTATACACATATGCACATACATGTCATCACATATATCGCA 380

Qy     304 ATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACA 351
             |  || |||  |||||||||||||||||||||| |||| |||||||||
Db     381 CTGATTATCTGGATTAAATTAAAACTAAAAATGCCTAACTTTCTAACA 428


RESULT 3
AAF81468/c
ID   AAF81468 standard; DNA; 1888 BP.
XX
AC   AAF81468;
XX
```

```
DT   08-JUN-2001  (first entry)
XX
DE   Corn promoter clone #700342364.
XX
KW   Corn; promoter; transgenic plant; herbicide resistance; ds.
XX
OS   Zea mays.
XX
PN   WO200119976-A2.
XX
PD   22-MAR-2001.
XX
PF   13-SEP-2000; 2000WO-US25078.
XX
PR   16-SEP-1999;   99US-0154182.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Anderson HM,  Chay CA,  Chen G,  Conner TW;
XX
DR   WPI; 2001-244796/25.
XX
PT   Novel promoter nucleic acid sequences useful for regulating
PT   heterologous gene expression in plants, comprising regulatory sequences
PT   located upstream to plant DNA structural coding sequences -
XX
PS   Claim 1; Page 95; 101pp; English.
XX
CC   The present invention relates to novel corn promoter sequences (see
CC   AAF81456-AAF81478). The promoter sequences are useful for conferring
CC   expression of a second polynucleotide molecule in a transgenic plant
CC   tissue. In addition, the promoter sequences are useful for providing
CC   plants with herbicide resistance. The promoter sequences are suitable for
CC   selectively modulating expression of any operatively linked gene and
CC   provide additional regulatory element diversity in a plant expression
CC   vector in gene stacking approaches. The present sequence is one such corn
CC   promoter sequence isolated in the present invention.
XX
SQ   Sequence 1888 BP; 569 A; 443 C; 423 G; 453 T; 0 other;

  Query Match             9.5%;  Score 124.6;  DB 22;  Length 1888;
  Best Local Similarity   63.9%;  Pred. No. 4.6e-22;
  Matches  227;  Conservative    0;  Mismatches  114;  Indels   14;  Gaps    2;

Qy      78 CCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGTACGGCTACATCGAA 137
           | ||| ||| |||||||  ||||| | ||  |||||||||| || ||||   || ||||||
Db     784 CTGTTTGAGGGACTGCACTGCGTAGACCTGCCTGCATCGAGTTCGTACAATTATATCGAA 725

Qy     138 CATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCC 197
           |||| |||||||||||||| |||||||| | | | || || | || |||||||||  || |
Db     724 CATACACACGAGATGTCTTGTGTGAATTGGGCCGCTGATACTTTGAGCATCGGTCCCTTC 665

Qy     198 GTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATAC 257
             |||| || | ||||| | | || || |      | |  || |||| ||||| ||||||
Db     664 ACAGGGAACTTCCTGTTTTCCATACTTTT--TGCGTGTCTTTTTTGTGTACTGCTTATAC 607

Qy     258 GAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAAT 317
           || ||||  |              |||||| || ||||||| || |||||  ||||||  |
Db     606 GAATAGTGGT-----------ATACATGTCGTCACATAAATTACAATGGTTTTCTGGAC 559

Qy     318 TAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCA 377
           ||||||||||||  |||| || | |||||||||      ||| |   |    ||    |
Db     558 TAAATTAAAACTAGTAATGCCTCATTTTCTAACATATTTCACAGTAATAAAAGTTAGGTA 499

Qy     378 ACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACAT 432
            || ||    || | |      | | ||  |     | | || |||  |  | ||
Db     498 TCAGCTAATTCTTTGGTGATAGGGTATAAATGACGATGCTCTATATTTATCAGAT 444


RESULT 4
AAS45444
ID   AAS45444 standard; DNA; 6327 BP.
XX
AC   AAS45444;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Chemically pretreated genomic DNA associated with cell cycle #75.
XX
KW   Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW   human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW   graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW   arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW   immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168911-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02945.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602751/68.
XX
PT   Designing primers and probes for analysing diseases associated with
PT   cytosine methylation state e.g. arthritis, cancer, aging,
PT   arteriosclerosis comprising fragments of chemically modified genes
PT   associated with cell cycle -
XX
PS   Claim 1; SEQ ID No 149; 28pp; English.
XX
CC   Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC   molecules associated with the cell cycle and specific PCR primers of the
CC   invention. The sequences are useful for detecting the methylation state
CC   of all CpG dinucleotides in a sequence and therefore for analysing
CC   associated diseases. By analysing cytosine methylations in the pretreated
CC   DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC   of existing diseases or the predisposition to specific diseases can be
CC   ascertained. The parameters may be compared to another set of genetic
CC   and/or epigenetic parameters, the differences serving as basis for
CC   diagnosis and/or prognosis events which are disadvantageous to patients.
CC   The sequences of the invention are useful for the diagnosis and therapy
CC   of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC   aging, glomerular disease, Lewy body disease, arthritis,
CC   arteriosclerosis, solid tumours and cancers.
XX
SQ   Sequence 6327 BP; 1833 A; 93 C; 1407 G; 2994 T; 0 other;

  Query Match             3.8%;  Score 50.2;  DB 22;  Length 6327;
  Best Local Similarity   60.7%;  Pred. No. 0.013;
  Matches   82;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     211 TGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACA 270
           | || ||  || ||  |  |   ||| ||   ||||     |||||    ||  |||| |
Db    2657 TATTGTTGATAGTTTGGTTTGAATTTTTTTAGGTTTTTGTTTTATATATATATATATATA 2716

Qy     271 TACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTA 330
           || ||  | ||| ||||    | ||||||    |||||||| |||| | ||||||||
Db    2717 TATATATATATATATATATATATATATATTTATATATTTTTGTAAAGTGAATTAAAAGAG 2776

Qy     331 AAAATGACTAAATTT 345
           | || | |||| ||
Db    2777 TATATAAATAAAGTT 2791
```

```
RESULT 5
ABL32458/c
ID   ABL32458 standard; DNA; 10133 BP.
XX
AC   ABL32458;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 431.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 431; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 10133 BP; 3075 A; 70 C; 2051 G; 4937 T; 0 other;

  Query Match             3.8%;  Score 50;  DB 24;  Length 10133;
  Best Local Similarity   51.0%;  Pred. No. 0.018;
  Matches  149;  Conservative    0;  Mismatches  135;  Indels    8;  Gaps    1;

Qy     219 TTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCA 278
           ||| |     ||| | | |      |  ||   | |||  | ||  |||  ||| || ||
Db    1416 TTAATCAAAAATACTCTAAACTAACTCAACAACTAATAAAACTACATATCTATATATACA 1357

Qy     279 CATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGAC 338
           |||||  ||  ||||||||   |  |||  | | ||||| | ||| | |  |      |
Db    1356 TATACAAATTCTCACATATTCAAATATACATATATAAATAATATTTACAAAATCCTATAT 1297

Qy     339 TAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTC----- 393
           ||||| | | | |  ||  | |    ||  |    | |    | | ||  |||||
Db    1296 TAAATATTTTAAAAAAAAAAAAAAAAAAATTCAAATACTCACATTCTAATTATTCAATAT 1237

Qy     394 ---TACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACA 450
             || ||| || || ||   |  ||  |  |||||  | |     ||||||  |||| |
Db    1236 TCTTAAATTATTATACTTAACAAATCTATAAATAAAACATAATACATACAATAAAAAAAA 1177

Qy     451 GTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATA 502
           | | ||  ||   ||||||    ||   |  |    || ||| |  | | |
Db    1176 CTTCATTTAACATTTATATATCAAATTCCTACCATATACCTAATAATAAAAA 1125


RESULT 6
ABL34124/c
ID   ABL34124 standard; DNA; 73334 BP.
XX
AC   ABL34124;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2097.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2097; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 73334 BP; 18968 A; 858 C; 15329 G; 38179 T; 0 other;

  Query Match             3.8%;  Score 49.4;  DB 24;  Length 73334;
  Best Local Similarity   53.3%;  Pred. No. 0.06;
  Matches  104;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy     252 TTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTT 311
           |||||    ||  |||| ||| ||  | ||| ||||    | ||||||  | |||| | |
Db   59235 TTATATATATATATATATATATATATATATATATATATATATATATATATATCTATATATAT 59176

Qy     312 CTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTT 371
           |||||  || ||||   | |||| | ||||| | || ||||||  ||
Db   59175 ATAAATATAAATAAATAAATAAATAAATAAATATTTACCACCAAAAACTAAAACTAAAAA 59116

Qy     372 TCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACA 431
             || ||    |  ||| | |        |   |  | | |||    | || || ||
```

```
Db  59115 CAATTAAAAAAACAAAATATACAAACACACATTCCATTAACTATCAAAAAAAAATCATCA 59056

Qy    432 TAGTCTACAATGGAA 446
          |  ||||  ||  ||
Db  59055 TCATCTATCATATAA 59041


RESULT 7
ABL34241/c
ID   ABL34241 standard; DNA; 11691 BP.
XX
AC   ABL34241;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2214.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 2214; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 11691 BP; 3695 A; 49 C; 1861 G; 6086 T; 0 other;

  Query Match             3.8%;  Score 49.2;  DB 24;  Length 11691;
  Best Local Similarity   51.1%;  Pred. No. 0.03;
  Matches  140;  Conservative    0;  Mismatches  133;  Indels    1;  Gaps    1;

Qy    160 TGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCT 219
          | || | |    ||| |  |||||  ||   | | |   ||   || ||    ||   |
Db   6938 TTAAAATATAACCTATTTACTTAAATATTAATAAATATATACACTAAATATATTTACCCA 6879

Qy    220 TATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 279
          || || |  ||| |     |    ||||   || |  | | ||  |||||| | |   |
Db   6878 TAATTCTAAATACTACATATACATTTTAAATATAACTC-ATTAAATATACACAAAAAAAT 6820

Qy    280 ATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACT 339
          ||| ||||   |||||| |  |  |||    | || ||  ||   | ||  | |||   |
Db   6819 ATATATATATACACATAAAATAACATAAAAATTTACATATAAAATACACATATAATATTT 6760

Qy    340 AAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATT 399
          |    ||||  |  ||  ||||  |||   ||| |  |  || | ||  ||   |  ||
Db   6759 ACTAATCTATAAATAAATACATATTAACACTTTATTTATTAAATATATATAAATTCTATA 6700

Qy    400 GTTCTATTTCGAATTTCACTCTATAAACAACATA 433
          |||  |     | || |||  || ||| || | |
Db   6699 TTTCAAAAATAATTTACACAATAAAAAAAATACA 6666


RESULT 8
AAZ17263
ID   AAZ17263 standard; cDNA; 1337 BP.
XX
AC   AAZ17263;
XX
DT   12-OCT-1999  (first entry)
XX
DE   Human gene expression product cDNA sequence SEQ ID NO:4735.
XX
KW   Human; gene; gene expression product; diagnosis; therapy; probe;
KW   detection; mapping; tissue typing; profiling; forensic; cancer;
KW   genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9938972-A2.
XX
PD   05-AUG-1999.
XX
PF   28-JAN-1999;   99WO-US01619.
XX
PR   03-APR-1998;   98US-0080666.
PR   28-JAN-1998;   98US-0072910.
PR   24-FEB-1998;   98US-0075954.
PR   31-MAR-1998;   98US-0080114.
PR   03-APR-1998;   98US-0080515.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Crkvenjakov R,  Dickson M,  Drmanac R,  Drmanac S;
PI   Escobedo J,  Garcia PD,  Garcia V,  Giese K,  Innis MA;
PI   Jones WL,  Kassam A,  Kennedy GC,  Kita D,  Labat I;
PI   Lamson G,  Leshkowitz D,  Pot D,  Randazzo F,  Reinhard C;
PI   Stache-Crain B,  Sudduth-Klinger J,  Williams LT;
XX
DR   WPI; 1999-494092/41.
XX
PT   Novel human genes and their expression products which are
PT   differentially expressed in different cell types
XX
PS   Claim 1; Page 2250-2251; 2479pp; English.
XX
CC   The present invention describes a library of human polynucleotides
CC   comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC   a method of detecting differentially expressed genes correlated with the
CC   cancerous state of a mammalian cell, comprising detecting at least one
CC   differentially expressed gene product in a test sample from a cell
CC   suspected of being cancerous, where the gene product is encoded by one
CC   of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC   polynucleotides can be used as a source of primers and probes, which can
CC   be used for a variety of purpose, e.g. detection of expression levels,
CC   mapping, tissue typing or profiling, forensics, genetic analysis and
CC   detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC   can be used for raising antibodies for experimental, diagnostic and
CC   therapeutic purposes. The polynucleotides may also be used to construct
CC   arrays for diagnostics (which may be used to determine function of an
CC   encoded protein); and to detect differences in expression levels between
CC   two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC   identify a genetic predisposition or susceptibility to a disease such as
```

```
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 1337 BP; 42 A; 577 C; 27 G; 22 T; 669 other;

  Query Match             3.7%;  Score 48.8;  DB 20;  Length 1337;
  Best Local Similarity   32.4%;  Pred. No. 0.015;
  Matches  101;  Conservative    0;  Mismatches  211;  Indels    0;  Gaps    0;

Qy     755 CACTACCCCAACCTCTCACCGAAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGC 814
           | |   |       |             |     |  |       |            |
Db     350 CCCNCNCNNNNNNNCNNNNNCNNCNNNNCNCCCNNNNCCNCNNNNNNCNCNNNNNNNNC 409

Qy     815 CCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCC 874
           |||   | |   | || |   | |        ||  |  |        |     ||   ||
Db     410 CCCCCNCNCCNNCNCCCCNNNCNCCCNNNNNNCCNCCCCCNCNNNNNCNNNNCCCCCCCC 469

Qy     875 GTGTCGCGGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCC 934
                 |  |      | |||||  ||| | | |    ||||  |||  ||   ||
Db     470 NNNNNNNNCNNCCCCNNCCNCNCCCCCNCCCCNCNCCCCCCCCCCCCCCCCCNCCCNNCCC 529

Qy     935 CGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGA 994
           |    |   || | | |  | |   |   || | |         |||     | |||
Db     530 CNNCCCNCCCCNCNCNCNCCCCCNNNCCCNNCCCCNCCNNNNCCCCCCNCCNCCCCCCCN 589

Qy     995 CACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCACGGCCGACTGC 1054
             | |   |  |||           |   ||     |   | |||| | |  |      |
Db     590 NCCNCNCNCCCCCCNCCCCCCCCCNCNNNCCCCCCNCCNCNCCCCCCCCCCCCCCNNNNCCC 649

Qy    1055 GCTGCGCCGCCC 1066
            |  | || |||
Db     650 CCCCCCCCCCCC 661


RESULT 9
AAH93294/c
ID   AAH93294 standard; DNA; 399 BP.
XX
AC   AAH93294;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Plasmodium falciparum MAL3P8 polynucleotide SEQ ID NO 16.
XX
KW   Human; antisense-therapy; gene-therapy; diagnostic; forensic;
KW   gene mapping; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200152616-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US35190.
XX
PR   23-DEC-1999;   99US-0471275.
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-451890/48.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
PS   Example 4; Page 106; 135pp; English.
XX
CC   The invention relates to an isolated human polynucleotide (AAH75398)
CC   encoding a novel polypeptide (AAG64527) useful in antisense-therapy and
CC   gene-therapy, in diagnostics, forensics, gene mapping and identification
CC   of mutations responsible for genetic disorders and other traits.
CC   Polynucleotide sequences with potential homology were also identified
CC   (AAH93283-AAH93356).
XX
SQ   Sequence 399 BP; 225 A; 24 C; 15 G; 135 T; 0 other;

  Query Match             3.7%;  Score 48;  DB 22;  Length 399;
  Best Local Similarity   47.1%;  Pred. No. 0.014;
  Matches  147;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy     208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
           || | || ||  ||| | ||   ||||||||  |||  ||     |  |    ||  | |
Db     320 TTTTTTTTTTTATATATTTGTCAATTTTTATGTTTGCATAAACGATTAATATATATCTTT 261

Qy     268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
             |||||    ||| |||| |  | ||||||    |||| | | || ||  |  || |
Db     260 TTTTACATTATAATATATATAAATATATATATATATATATATATATATATATATATATAT 201

Qy     328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
           || | ||   |  |||| |   || | |   |  |  | |||| |             |
Db     200 ATATATATTTATTTATTTATTTTACGATAGTTTTAATTTTTTTTTTTTTATCTGCGAATTA 141

Qy     388 CTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAA 447
             | |  ||| |  | |||  | |     | |||||   | ||  | || ||   |||
Db     140 AACATTTTAATTATGTTCTTTATATTAATTATATATAATATATTATATAAAAATAAAA 81

Qy     448 ACAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAGTCGT 507
             ||    || |   | | |  | |   | | |    |       | | |||  ||
Db      80 TTAGAATGTTTTGATATTTTTAAGGTTTTTTTTTTTTTTTTTTTGATACACAATTCATTA 21

Qy     508 TTGAAGATTGAA 519
           || ||  |  ||
Db      20 TTTAAATTATAA 9


RESULT 10
AAS56505/c
ID   AAS56505 standard; cDNA; 326 BP.
XX
AC   AAS56505;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human cDNA for an ovarian cancer protein #129.
XX
KW   Human; ss; ovarian cancer protein; cancer; tumour; ovarian cancer;
KW   endometrial cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200170976-A2.
XX
PD   27-SEP-2001.
XX
PF   20-MAR-2001; 2001WO-US09062.
XX
PR   21-MAR-2000; 2000US-190710P.
PR   22-JUN-2000; 2000US-213748P.
PR   19-DEC-2000; 2000US-257276P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Pyle RA,  Stolk JA;
XX
DR   WPI; 2001-607531/69.
XX
PT   Nucleic acids encoding 222 polypeptides associated with ovarian and
PT   endometrial cancers, useful for diagnosing, preventing and treating
PT   cancers -
XX
PS   Claim 1; Page 157; 187pp; English.
```

```
XX
CC   The invention relates to human polynucleotides encoding proteins
CC   associated with ovarian and endometrial cancers. The polynucleotides and
CC   the proteins they encode may be used in the prevention, diagnosis and
CC   treatment of diseases associated with the inappropriate expression of
CC   ovarian and endometrial cancer polypeptides (OECPs). For example, the
CC   polynucleotide (or an expression vector comprising the polynucleotide)
CC   and the OECP may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of OECPs by expressing inactive proteins or to
CC   supplement the patients own production of them. Additionally, the
CC   polynucleotide may be used to produce the OECPs, by inserting the nucleic
CC   acids into a host cell and culturing the cell to express the protein. The
CC   polynucleotide and its complementary sequences may also be used as DNA
CC   probes in diagnostic assays to detect and quantitate the presence of
CC   similar nucleic acids in samples, and therefore which patients may be in
CC   need of restorative therapy. The OECPs may also be used as antigens in
CC   the production of anti-OECP antibodies and in assays to identify
CC   modulators of it's expression and activity. The anti-OECP antibodies and
CC   antagonists may also be used to down regulate expression and activity.
CC   The anti-OECP antibodies may also be used as diagnostic agents for
CC   detecting the presence of OECP in samples (e.g. by enzyme linked
CC   immunosorbant assay (ELISA)) and hence diagnose patients with
CC   cancers. The present sequence is a ovarian and endometrial cancer linked
CC   cDNA of the invention.
XX
SQ   Sequence 326 BP; 130 A; 48 C; 34 G; 114 T; 0 other;

  Query Match             3.6%;  Score 47.8;  DB 22;  Length 326;
  Best Local Similarity   55.0%;  Pred. No. 0.014;
  Matches   94;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     211 TGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACA 270
           | || || || ||| |   | |||      |||   ||   ||  |  | |||  |||||
Db     326 TTTTTTTTTTTTTTTTTTTTTTTTCAGGGATTGAACACATTTTTAATTATTAGAAATACA 267

Qy     271 TACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTA 330
           ||||||||| |||  || | |   || |||    | |  |   | |     || ||||  |
Db     266 TACATGCATATAATTAACCTTTAATTTATTTTTAAAAATAAATTATGGTCATGAAAAACA 207

Qy     331 AAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAA 381
           | | ||  |||| ||  ||   ||  |||||   | | | ||| | || ||
Db     206 ATAGTGCTTAAAATTGAAATTTCATAGACATGTAATTTTATTCCCTAAAAA 156


RESULT 11
ABL34125/c
ID   ABL34125 standard; DNA; 73334 BP.
XX
AC   ABL34125;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2098.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2098; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 73334 BP; 24605 A; 858 C; 14432 G; 33439 T; 0 other;

  Query Match             3.6%;  Score 47.8;  DB 24;  Length 73334;
  Best Local Similarity   51.1%;  Pred. No. 0.16;
  Matches  112;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy     232 TTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATC 291
           | |||||   | || | | ||||    | ||   | | |||||| || | | |   | |
Db   66278 TATTTATAAATCTTAAATAATTAACTTAATATAAAAAAATACATACAAAAAAAACACCTT 66219

Qy     292 ACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACA 351
           | | || | ||||||||||  ||||   |   |||   |||||  |  | | ||| || |
Db   66218 AAAAATTTTACAATATTTTAATAAAAATACAAAAATAAAAAAAAAAAAACAATTCAAATA 66159

Qy     352 CCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGA 411
           |||  | |   ||||   || | |    | |       | | || |       |     |
Db   66158 ACAAACAAAACATAATAACTTATACCTATAATCCCAACACTTTAAAAAACGAAAACGAAA 66099

Qy     412 ATTTCACTCTATAAACAACATAGTCTACAATGGAAAACA 450
           |  |||||  ||    || ||      | |  | |||||
Db   66098 AAATCACTTAATCCCAAAAATTCGAATCCAACGTAAACA 66060


RESULT 12
ABL33574/c
ID   ABL33574 standard; DNA; 6627 BP.
XX
AC   ABL33574;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 1547.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
```

PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 1547; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 6627 BP; 2052 A; 117 C; 1386 G; 3072 T; 0 other;

Query Match 3.6%; Score 47.4; DB 24; Length 6627;
Best Local Similarity 49.2%; Pred. No. 0.069;
Matches 155; Conservative 0; Mismatches 156; Indels 4; Gaps 1;

```
Qy   192 TACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGA 251
         |||||  ||   ||| | ||  |  ||   |   | |||  |   |||     | | | |
Db  6510 TACTCTTTAAAATACCTACTAATTCTCAATTCCTTTCATCCTAACATTAACCCTAAATAA 6451

Qy   252 TTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTT 311
         ||| |    || | ||  | | ||  | | ||  |  | || | | || | |  | |||
Db  6450 TTACAATTATATTAATTTACAAATAAAAAAACTAAAAACCAAAAAAATTAAATAACTTTA 6391

Qy   312 CTAAATT----AAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAAT 367
         |||| ||     || |||||   ||||  |||||| ||  ||| | ||  ||       ||
Db  6390 CTAACTTTACCCAACTAAAAAATAAAAAAACTAAAATTTAAACCCAAAAAACTAAACCAT 6331

Qy   368 GTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAAC 427
          |  |||  || | ||  ||    |  ||||  |  |  |  | | ||| |  ||
Db  6330 ATAATCTAAAAAACCTAAACTCCATAATCATTACTAAACATAAACTATCAATTAATCCTT 6271

Qy   428 AACATAGTCTACAATGGAAAACAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACA 487
         |  | |       ||  | ||||| | || | | |   |  | |  | |          |||
Db  6270 ATAACAACTCCAAAACGCAAACAATACTATCTCCATTTCAAAAACCAACAAACTAAAACA 6211

Qy   488 ACATAAGACAATATA 502
          ||   || ||  ||
Db  6210 CCAATCGAAAAACTA 6196
```

RESULT 13
AAA70152/c
ID AAA70152 standard; DNA; 6033 BP.
XX
AC AAA70152;
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:285.
XX
KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PF 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
PS Disclosure; Page 493-495; 577pp; English.
XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic lifecycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 6033 BP; 3019 A; 437 C; 707 G; 1870 T; 0 other;

Query Match 3.6%; Score 47.2; DB 21; Length 6033;
Best Local Similarity 49.6%; Pred. No. 0.074;
Matches 121; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

```
Qy   205 ACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGT 264
         ||| ||| ||||   ||       || | | || | ||  |||   ||  ||    || |
Db  2523 ACAATCTTTTCTATGTACAACCATATTTATATAATATTTATTATATATATTAGTTATACT 2464

Qy   265 TATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTA 324
           || ||  ||    ||     | ||| ||  | |||   || ||||   ||| |||||  |
Db  2463 GTTAAATGTATTATCAATATTTTCAACATTTTTATTTAAAAATTTAAATAATTTAAAAAA 2404

Qy   325 AAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTT 384
              |||| | | | ||| | ||   |  |    |||| ||  | || | ||    |
Db  2403 TTTTGAAAATTCATTCAATATATATTTCATATATAATTGAAAGATATTATACACATTCCA 2344

Qy   385 TACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGG 444
         | | | |  | | || |  | |||     ||   | | | |       ||    | ||| |
Db  2343 TTCTTCTAATCCTTTTTCATTTTTTTGTCTTTTTTTTTTTTATATTATTATAAAAAAATTG 2284

Qy   445 AAAA 448
          |||
Db  2283 GAAA 2280
```

RESULT 14
AAL16482
ID AAL16482 standard; cDNA; 281 BP.
XX

AC AAL16482;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 8939.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 1617; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 281 BP; 93 A; 32 C; 39 G; 117 T; 0 other;

Query Match 3.6%; Score 46.8; DB 22; Length 281;
Best Local Similarity 54.7%; Pred. No. 0.024;
Matches 93; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 212 GTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACAT 271
Db 9 GTTTTTTTTTTTTTTTTTTTTTTCAGGGTTTGAACACATTTTTAATTATTAGAAATACAT 68

Qy 272 ACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAA 331
Db 69 ACATGCTTATAATTAACCTTAAATTTATTTTTAAAAATAAATTATGGTCATGAAAAACAA 128

Qy 332 AAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAA 381
Db 129 TATGGCTTAAAATTGAAATTTCATAGACATGTAATTTTATTCCCTAAAAA 178

RESULT 15
AAS61315/c
ID AAS61315 standard; DNA; 6107 BP.
XX
AC AAS61315;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #270.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
PS Disclosure; SEQ ID No 276; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases, by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6107 BP; 1822 A; 187 C; 1337 G; 2761 T; 0 other;

Query Match 3.6%; Score 46.8; DB 24; Length 6107;
Best Local Similarity 49.2%; Pred. No. 0.094;
Matches 123; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 220 TATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 279
Db 2615 TTTATATATATATATATTTATATATATATTTATTACATTTTTATTTCCACAAACTCTATC 2556

Qy 280 ATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACT 339
Db 2555 ATAAATCTAATCATTAATAACATATATTTAATTAAAAAAATATATAACTTAATAAATTAT 2496

Qy 340 AAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATT 399
Db 2495 ATAAATATTATATTAAAATAAATTTCACATTATCTTAATTTACAAAAACTAAAATATCTA 2436

```
Qy     400 GTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGT 459
               |    ||| ||| || | |||| |      |  |||||   | | |   | |
Db    2435 AAAAACTAAAAAATATCAATCAACAAACTATTATTTTAACAATTATACATAACACATAAA 2376

Qy     460 ACGACTATAT 469
           |   ||| ||
Db    2375 AATTCTAAAT 2366
```

Search completed: November 7, 2002, 10:47:44
Job time : 320.564 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:44:03 ; Search time 39.0149 Seconds
(without alignments)
8247.617 Million cell updates/sec

Title: US-09-905-558C-3
Perfect score: 1310
Sequence: 1 cccatcgctgctttgtctac.....caagggaaagtgatcggatg 1310

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	51.2	3.9	7218	1	US-08-232-463-14	Sequence 14, Appl
	2	45.2	3.5	4257	2	US-08-690-473-1	Sequence 1, Appli
	3	45.2	3.5	4257	4	US-09-259-821A-1	Sequence 1, Appli
	4	45.2	3.5	4257	4	US-08-843-659-1	Sequence 1, Appli
c	5	45.2	3.5	12001	1	US-08-458-568A-11	Sequence 11, Appl
c	6	43.4	3.3	2781	3	US-08-749-522-4	Sequence 4, Appli
	7	43	3.3	796	4	US-09-007-119-14	Sequence 14, Appl
c	8	43	3.3	809	4	US-09-007-119-10	Sequence 10, Appl
c	9	41.8	3.2	1890	6	5312912-3	Patent No. 5312912
c	10	41.6	3.2	1441	4	US-08-821-994-63	Sequence 63, Appl
c	11	41	3.1	4254	2	US-08-443-639-7	Sequence 7, Appli
c	12	40.6	3.1	6152	4	US-08-973-462-1	Sequence 1, Appli
	13	40.4	3.1	1000	1	US-08-599-252-96	Sequence 96, Appl
	14	40.4	3.1	1000	5	PCT-US96-06352-96	Sequence 96, Appl
	15	40.4	3.1	1000	5	PCT-US96-06583-96	Sequence 96, Appl
c	16	40.4	3.1	10342	4	US-08-972-927-5	Sequence 5, Appli
	17	40	3.1	4673	1	US-07-638-431-1	Sequence 1, Appli
	18	40	3.1	4673	5	PCT-US92-00018-1	Sequence 1, Appli
c	19	40	3.1	5829	4	US-09-004-838-109	Sequence 109, App
c	20	39.6	3.0	51952	3	US-08-947-823-1	Sequence 1, Appli
c	21	39.2	3.0	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	22	39	3.0	320	4	US-09-165-264-11	Sequence 11, Appl
c	23	38.8	3.0	923	4	US-09-004-731-66	Sequence 66, Appl
c	24	38.8	3.0	923	4	US-08-749-699-66	Sequence 66, Appl
	25	38.8	3.0	152331	3	US-09-128-155-16	Sequence 16, Appl
c	26	38.6	2.9	320	4	US-09-165-264-13	Sequence 13, Appl
c	27	38.4	2.9	1368	3	US-08-874-563-5	Sequence 5, Appli
c	28	38.4	2.9	1368	3	US-08-577-483-14	Sequence 14, Appl
c	29	38.4	2.9	1736	3	US-09-182-816-22	Sequence 22, Appl
	30	38.4	2.9	1736	3	US-09-182-816-24	Sequence 24, Appl
c	31	38.4	2.9	1736	3	US-09-471-528-22	Sequence 22, Appl
	32	38.4	2.9	1736	3	US-09-471-528-24	Sequence 24, Appl
c	33	38.4	2.9	1736	4	US-09-634-530-22	Sequence 22, Appl
	34	38.4	2.9	1736	4	US-09-634-530-24	Sequence 24, Appl
c	35	38.4	2.9	4253	3	US-08-577-483-7	Sequence 7, Appli
	36	38.2	2.9	4526	1	US-07-855-412B-4	Sequence 4, Appli
	37	38.2	2.9	4526	2	US-08-308-887A-4	Sequence 4, Appli
	38	38.2	2.9	4526	3	US-08-881-094-4	Sequence 4, Appli
c	39	38	2.9	319	4	US-09-165-264-8	Sequence 8, Appli
	40	37.8	2.9	611	4	US-09-385-982-393	Sequence 393, App
c	41	37.8	2.9	899	1	US-07-820-154A-3	Sequence 3, Appli
c	42	37.8	2.9	899	2	US-08-097-554A-3	Sequence 3, Appli
c	43	37.8	2.9	899	3	US-08-480-640A-3	Sequence 3, Appli
c	44	37.8	2.9	899	3	US-08-295-802-3	Sequence 3, Appli
c	45	37.8	2.9	899	4	US-08-488-237A-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14

  Query Match             3.9%;  Score 51.2;  DB 1;  Length 7218;
```

```
  Best Local Similarity   4.3%;  Pred. No. 0.00017;
  Matches   17;  Conservative  219;  Mismatches  162;  Indels    0;  Gaps    0;

Qy     792 CGGCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTC 851
           | |||  || |  ||| | | |  :: : ::: :::::: : ::: :  : :: ::::::
Db    1043 CTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1102

Qy     852 TGCAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAACCACCCCCAACCCACTCA 911
           : :  :: ::: ::  :: : :: : :: :    :  : :  :: ::::: ::: :::
Db    1103 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1162

Qy     912 CGAAACCCCGTCCCGGCCGTGCCCGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGC 971
           :    :::: ::::  :: : ::: : ::  ::: : :::  :  :    :  ::: : :
Db    1163 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1222

Qy     972 TGCTGAGTCCCCTGGACACCCGACACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCT 1031
           : ::   ::::::   : :::  : :::: ::  :::::: ::: ::: ::::    :::
Db    1223 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1282

Qy    1032 CATCTGCCCCCACGGCCGACTGCGCTGCGCCGCCCGGATATATATACCCATCGTTATCGA 1091
           : ::: ::::: :  ::  :: : :: : :: :::   : : : : ::: :: :: ::
Db    1283 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1342

Qy    1092 TCGATCGATCGCGTCACTCACGGGTAGCTCATGGTCGAGCGTAGCATGCAGGAACTTATT 1151
           ::  ::  :: : :: ::: :   :  ::: :  ::   : :  : : :    ::: ::
Db    1343 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1402

Qy    1152 TGCCGTGCGCTCCCAGGTCTCCGCTCGCGTGCCTTCCA 1189
           : :: : : :::::   ::::: ::: : : : | |||
Db    1403 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCA 1440


RESULT 2
US-08-690-473-1
; Sequence 1, Application US/08690473
; Patent No. 5876923
;  GENERAL INFORMATION:
;    APPLICANT:  Leopardi, Rosario
;    APPLICANT:  Roizman, Bernard
;    TITLE OF INVENTION:  HERPES SIMPLEX VIRUS ICP4 AS AN
;    TITLE OF INVENTION:  INHIBITOR OF APOPTOSIS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/690,473
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  ARCD:239
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4257 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-690-473-1

  Query Match             3.5%;  Score 45.2;  DB 2;  Length 4257;
  Best Local Similarity   49.3%;  Pred. No. 0.0063;
  Matches  145;  Conservative    0;  Mismatches  148;  Indels    1;  Gaps    1;

Qy     714 CAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACCCCAACCTCTCAC 773
           | | |||    || || |||| | |  ||  || | |  | | |  |||   ||   | |
Db    2513 CCGCCTCCGCCGCGCCGCGGGAGGGGCGCAAGCGCAAGAGTCCCGGCCCGGCCCGGCCGC 2572

Qy     774 CGAAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACG 833
           |   |  | |||          |     | |||| |  | ||  || |  | | | |  |
Db    2573 CCGGAGGCGGCGGCCCGCGACCCCCGAAGACGAAGAAGAGCGGCGCGGACGCCCCCGGCT 2632

Qy     834 TCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAA 893
             ||| |  |  ||| || |   ||| | |   ||  ||  | | | |     |
Db    2633 CGGACGCCCGCGCCCCCCTCCCCGCGCCCGCGCCCCCCTCCACGCCCCCGGGGCCCGAGC 2692

Qy     894 CCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCGTGTCGGTCCGCGCTCGG 953
           || |||||  ||  | | ||   |||||  ||| |   || |  | | | ||||||  | |
Db    2693 CCGCCCCCGCCCAGCCCGCGGCGCCCCGGGCCGCCGCGGCGCAGGCCCGCCCGCGCCCCG 2752

Qy     954 CAA-CGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTGTCGGC 1006
               || | ||  |||  | || ||| |||  | | || |     ||| ||||
Db    2753 TGGCCGTGTCGCGCCGGCCCGCCGAGGGCCCCGACCCCCTGGGCGGCTGGCGGC 2806


RESULT 3
US-09-259-821A-1
; Sequence 1, Application US/09259821A
; Patent No. 6210926
; GENERAL INFORMATION:
;  APPLICANT:  LEOPARDI, ROSARIO
;  APPLICANT:  ROIZMAN, BERNARD
;  TITLE OF INVENTION:  HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
;  FILE REFERENCE:  ARCD:317
;  CURRENT APPLICATION NUMBER: US/09/259,821A
;  CURRENT FILING DATE:  1999-03-01
;  PRIOR APPLICATION NUMBER:  08/690,473
;  PRIOR FILING DATE:  1996-07-26
;  NUMBER OF SEQ ID NOS:  2
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4257
;   TYPE: DNA
;   ORGANISM: HERPES VIRUS, TYPE 1
US-09-259-821A-1

  Query Match             3.5%;  Score 45.2;  DB 4;  Length 4257;
  Best Local Similarity   49.3%;  Pred. No. 0.0063;
  Matches  145;  Conservative    0;  Mismatches  148;  Indels    1;  Gaps    1;

Qy     714 CAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACCCCAACCTCTCAC 773
           | | |||    || || |||| | |  ||  || | |  | | |  |||   ||   | |
Db    2513 CCGCCTCCGCCGCGCCGCGGGAGGGGCGCAAGCGCAAGAGTCCCGGCCCGGCCCGGCCGC 2572

Qy     774 CGAAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACG 833
           |   |  | |||          |     | |||| |  | ||  || |  | | | |  |
Db    2573 CCGGAGGCGGCGGCCCGCGACCCCCGAAGACGAAGAAGAGCGGCGCGGACGCCCCCGGCT 2632

Qy     834 TCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAA 893
             ||| |  |  ||| || |   ||| | |   ||  ||  | | | |     |
Db    2633 CGGACGCCCGCGCCCCCCTCCCCGCGCCCGCGCCCCCCTCCACGCCCCCGGGGCCCGAGC 2692

Qy     894 CCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCGTGTCGGTCCGCGCTCGG 953
           || |||||  ||  | | ||   |||||  ||| |   || |  | | | ||||||  | |
Db    2693 CCGCCCCCGCCCAGCCCGCGGCGCCCCGGGCCGCCGCGGCGCAGGCCCGCCCGCGCCCCG 2752

Qy     954 CAA-CGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTGTCGGC 1006
               || | ||  |||  | || ||| |||  | | || |     ||| ||||
Db    2753 TGGCCGTGTCGCGCCGGCCCGCCGAGGGCCCCGACCCCCTGGGCGGCTGGCGGC 2806


RESULT 4
US-08-843-659-1
```

```
; Sequence 1, Application US/08843659
; Patent No. 6218103
;  GENERAL INFORMATION:
;    APPLICANT:  Leopardi, Roasrio
;    APPLICANT:  Roizman, Bernard
;    TITLE OF INVENTION:  HERPES SIMPLEX VIRUS US3 AND ICP4 AS
;    TITLE OF INVENTION:  INHIBITORS OF APOPTOSIS
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  United States
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/843,659
;      FILING DATE:  Concurrently Herewith
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  ARSB:519
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (512) 418-3000
;      TELEFAX:  (512) 474-7577
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4257 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-843-659-1

  Query Match             3.5%;  Score 45.2;  DB 4;  Length 4257;
  Best Local Similarity   49.3%;  Pred. No. 0.0063;
  Matches  145;  Conservative    0;  Mismatches  148;  Indels    1;  Gaps    1;

Qy     714 CAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACCCCAACCTCTCAC 773
           | | |||    || || |||| | |  ||  || | |  | | |  |||   ||   | |
Db    2513 CCGCCTCCGCCGCGCCGCGGGAGGGGCGCAAGCGCAAGAGTCCCGGCCCGGCCCGGCCGC 2572

Qy     774 CGAAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACG 833
           |   | | |||          |     | |||| |  | ||  || |  | | | |  |
Db    2573 CCGGAGGCGGCGGCCCGCGACCCCCGAAGACGAAGAAGAGCGGCGCGGACGCCCCCGGCT 2632

Qy     834 TCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAA 893
           |||  |  |  ||| || |   ||| | |   ||   ||  | | | |     |
Db    2633 CGGACGCCCGCGCCCCCCTCCCCGCGCCCGCGCCCCCCTCCACGCCCCCGGGGCCCGAGC 2692

Qy     894 CCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCCGTGTCGGTCCGCGCTCGG 953
           || |||||  ||  | | ||   |||||  ||| |   || |  | | | ||||||  | |
Db    2693 CCGCCCCCGCCCAGCCCGCGGCGCCCCGGGCCGCCGCGGCGCAGGCCCGCCCGCGCCCCG 2752

Qy     954 CAA-CGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTGTCGGC 1006
               || | ||  |||  | || |||  ||| |  | || |     ||| ||||
Db    2753 TGGCCGTGTCGCGCCGGCCCGCCGAGGGCCCCGACCCCCTGGGCGGCTGGCGGC 2806


RESULT 5
US-08-458-568A-11/c
; Sequence 11, Application US/08458568A
; Patent No. 5821339
;  GENERAL INFORMATION:
;    APPLICANT:  Schaffer, Priscilla A.
;    APPLICANT:  Yeh, Lily
;    TITLE OF INVENTION:  Compositions and Methods for Treatment of Herpesvirus
;    TITLE OF INVENTION:  Infections
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,568A
;      FILING DATE:  02-JUNE-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/065,146
;      FILING DATE:  05-MAY-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leary Ph.D., Kathryn R.
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  DFCI-0029
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Herpes simplex virus
;      STRAIN:  Herpes Simplex Virus Type 1
US-08-458-568A-11

  Query Match             3.5%;  Score 45.2;  DB 1;  Length 12001;
  Best Local Similarity   49.3%;  Pred. No. 0.011;
  Matches  145;  Conservative    0;  Mismatches  148;  Indels    1;  Gaps    1;

Qy     714 CAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACCCCAACCTCTCAC 773
           | | |||    || || |||| | |  ||  || | |  | | |  |||   ||   | |
Db    4161 CCGCCTCCGCCGCGCCGCGGGAGGGGCGCAAGCGCAAGAGTCCCGGCCCGGCCCGGCCGC 4102

Qy     774 CGAAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACG 833
           |   | | |||          |     | |||| |  | ||  || |  | | | |  |
Db    4101 CCGGAGGCGGCGGCCCGCGACCCCCGAAGACGAAGAAGAGCGGCGCGGACGCCCCCGGCT 4042

Qy     834 TCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAA 893
           |||  |  |  ||| || |   ||| | |   ||   ||  | | | |     |
Db    4041 CGGACGCCCGCGCCCCCCTCCCCGCGCCCGCGCCCCCCTCCACGCCCCCGGGGCCCGAGC 3982

Qy     894 CCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCCGTGTCGGTCCGCGCTCGG 953
           || |||||  ||  | | ||   |||||  ||| |   || |  | | | ||||||  | |
Db    3981 CCGCCCCCGCCCAGCCCGCGGCGCCCCGGGCCGCCGCGGCGCAGGCCCGCCCGCGCCCCG 3922

Qy     954 CAA-CGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTGTCGGC 1006
               || | ||  |||  | || |||  ||| |  | || |     ||| ||||
Db    3921 TGGCCGTGTCGCGCCGGCCCGCCGAGGGCCCCGACCCCCTGGGCGGCTGGCGGC 3868


RESULT 6
US-08-749-522-4/c
; Sequence 4, Application US/08749522
; Patent No. 6096950
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal
```

```
;    TITLE OF INVENTION:  FIBER-SPECIFIC PROMOTERS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  WI
;      COUNTRY:  U.S.A.
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/749,522
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baker, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  670513.90244
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5709
;      TELEFAX:  (414) 271-3552
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2781 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-749-522-4

  Query Match             3.3%;  Score 43.4;  DB 3;  Length 2781;
  Best Local Similarity   50.7%;  Pred. No. 0.016;
  Matches  104;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     228 CATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATAT 287
           ||   ||| || ||| |||| |    || |||    ||||| ||   |  |  || ||||
Db    1343 CAGGATTTAATAGTTATTTAGTTTAGATTCGAAATTTTATATATTATTTTATTTATATAT 1284

Qy     288 CATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCT 347
             | | |||||| | |||  | ||  |||| |  |  |   || ||||    | || |
Db    1283 TTTAATATATATTATAATTATATTTAAAATAATTTATATTTTATAAATTTTGATATATTA 1224

Qy     348 AACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATT 407
           |  |   |  | ||   ||| ||   |        |||   |||| ||      ||   |
Db    1223 ATAATATATTATATATAAATTTTAATTTATTTTTATTTTTATATTATATTAATATCATAT 1164

Qy     408 TCGAATTTCACTCTATAAACAACAT 432
           |  |||||   |   || |  | ||
Db    1163 TTTAATTTATTTTATTATATCATAT 1139


RESULT 7
US-09-007-119-14
; Sequence 14, Application US/09007119C
; Patent No. 6300541
; GENERAL INFORMATION:
;  APPLICANT: Lighfoot, David A.
;  APPLICANT: Gibson, Paul T.
;  APPLICANT: Merkem, Khalid
;  TITLE OF INVENTION: Soybean Sudden Death Syndrome Resistant Soybeans,
;  TITLE OF INVENTION: Soybean Cyst Nematode Resistant Soybeans and Methods of
;  TITLE OF INVENTION: Breeding and Identifying Resistant Plants
;  FILE REFERENCE: Sou Illinois 1268/2 Sequence Listing
; Patent No. 6300541
;  CURRENT APPLICATION NUMBER: US/09/007,119C
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,335
;  EARLIER FILING DATE: 1997-01-14
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;  LENGTH: 796
;  TYPE: DNA
;  ORGANISM: Glycine max
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (3)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (20)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (473)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (479)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (610)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (668)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (677)..(678)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (710)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (723)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (771)..(772)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (787)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (792)
;  OTHER INFORMATION: (a or c or g or t/u)
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: (796)
;  OTHER INFORMATION: (a or c or g or t/u)
US-09-007-119-14

  Query Match             3.3%;  Score 43;  DB 4;  Length 796;
  Best Local Similarity   54.8%;  Pred. No. 0.011;
  Matches   85;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     221 ATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACA 280
           |||||||  |  ||  || |    ||       ||||  || |   |||  ||  | ||||
Db     148 ATTTGTGAAATTTGGTACTCCCTTTGGACTCGTATATAAGAAAAAATAACTAATTTCACA 207

Qy     281 TACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTA 340
           |  ||       |  ||  |||||  ||||   |  | |||  | |||  ||||||  |
Db     208 TTAATTAAGAAAGTTAGTTCACATCATTTAATTTTACTAATCTCAAATAAAAAATAAGGT 267

Qy     341 AATTTCTAACACCAACGACATTGTAATGTTTTCTC 375
            ||||||||  |    |  |||||  ||  |  || ||
```

```
Db    268 TATTTCTAAAATGTCCTTCATTGGAACTTGTTATC 302


RESULT 8
US-09-007-119-10/c
; Sequence 10, Application US/09007119C
; Patent No. 6300541
; GENERAL INFORMATION:
;  APPLICANT: Lighfoot, David A.
;  APPLICANT: Gibson, Paul T.
;  APPLICANT: Merkem, Khalid
;  TITLE OF INVENTION: Soybean Sudden Death Syndrome Resistant Soybeans,
;  TITLE OF INVENTION: Soybean Cyst Nematode Resistant Soybeans and Methods of
;  TITLE OF INVENTION: Breeding and Identifying Resistant Plants
;  FILE REFERENCE: Sou Illinois 1268/2 Sequence Listing
; Patent No. 6300541
;  CURRENT APPLICATION NUMBER: US/09/007,119C
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,335
;  EARLIER FILING DATE: 1997-01-14
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 809
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(6)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (506)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (619)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (662)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (687)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (747)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (766)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (769)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (785)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (795)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (797)
;   OTHER INFORMATION: (a or c or g or t/u)
US-09-007-119-10

  Query Match             3.3%;  Score 43;  DB 4;  Length 809;
  Best Local Similarity   54.8%;  Pred. No. 0.011;
  Matches   85;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     221 ATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACA 280
           ||||||| |  ||  || |    ||        ||||  || |   |||  ||  | ||||
Db     457 ATTTGTGAAATTTGGTACTCCCTTTGGACTCGTATATAAGAAAAAATAACTAATTTCACA 398

Qy     281 TACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTA 340
           |  ||      |  ||  |||||  ||||   |  | |||  | |||  ||||||
Db     397 TTAATTAAGAAAGTTAGTTCACATCATTTAATTTTACTAATCTCAAATAAAAAATAAGGT 338

Qy     341 AATTTCTAACACCAACGACATTGTAATGTTTTCTC 375
           ||||||||  |    |  ||||| ||  | || ||
Db     337 TATTTCTAAAATGTCCTTCATTGGAACTTGTTATC 303


RESULT 9
5312912-3/c
;Patent No. 5312912
;    APPLICANT: HADWIGER, LEE A.;CHIANG, CHIN C.;HOROVITZ,
;DANIEL A.
;    TITLE OF INVENTION: PROCEDURES AND REGULATORY DNA SEQUENCES
;FOR GENETICALLY ENGINEERING DISEASE RESISTANCE AND OTHER
;INDUCIBLE TRAITS IN PLANTS
;    NUMBER OF SEQUENCES: 9
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/393,301
;      FILING DATE: 13-JUN-1989
;SEQ ID NO:3:
;      LENGTH: 1890
5312912-3

  Query Match             3.2%;  Score 41.8;  DB 6;  Length 1890;
  Best Local Similarity   53.3%;  Pred. No. 0.038;
  Matches   88;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy     232 TTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATC 291
           ||| | || |   ||       | || |  |   ||| ||| ||  | ||| ||||
Db     288 TTTATTTTTTAAATTTTCACACACACAAAAAAAAATATATATATATATATATATATAT 229

Qy     292 ACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACA 351
           | ||||||    |||| | | || ||  |  || |  || | || |  || |   || |
Db     228 ATATATATATATATATATATATATATATATATATATATATATATTAACAACTCAATATCT 169

Qy     352 CCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTAC 396
           |   |   ||| |  | |||| |  ||| |||| |||  |   ||
Db     168 CTGTCATAATTATTTTCTTTTTTTTAACTACTTAACCCTTATGAC 124


RESULT 10
US-08-821-994-63/c
; Sequence 63, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
;  APPLICANT: Greenland, Andrew J
;  APPLICANT: Thomas, Didier RP
;  APPLICANT: Jepson, Ian
;  TITLE OF INVENTION: Promoters
;  FILE REFERENCE: PPD 50108
;  CURRENT APPLICATION NUMBER: US/08/821,994A
;  CURRENT FILING DATE: 1997-03-22
;  EARLIER APPLICATION NUMBER: PCT/GB97/00729
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: GB 9606062.9
;  EARLIER FILING DATE: 1996-03-22
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
;   LENGTH: 1441
;   TYPE: DNA
;   ORGANISM: Brassica napus
US-08-821-994-63
```

```
  Query Match             3.2%;  Score 41.6;  DB 4;  Length 1441;
  Best Local Similarity   48.7%;  Pred. No. 0.037;
  Matches  113;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy     211 TGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACA 270
           | || || || ||| |   | ||||| || || |||  |  || |     |  || |
Db    1441 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1382

Qy     271 TACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTA 330
           |   |  | | | |||||||      |||     | ||  |  | | | ||   ||| |
Db    1381 TTTTTTGAGAAATATATCATTTTTATTATATTTCTCTTAATACATAATCAAGCCAAAGCA 1322

Qy     331 AAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTA 390
           || ||   |  | |   |  |        ||   ||  | ||| | ||       | ||||
Db    1321 AATATTCTTTGAATGAAATGATATTGACCAACATAGAGATTTATACAGATTAAATCCCTA 1262

Qy     391 TTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAAT 442
           |  |  ||   | | |||||||  |  |  ||| | |  ||| |  |  ||
Db    1261 TGGTGAATCAATTTTTTTCGAAATACTATACATACAAACAATAATACAAGAT 1210


RESULT 11
US-08-443-639-7/c
; Sequence 7, Application US/08443639
; Patent No. 5981843
;  GENERAL INFORMATION:
;    APPLICANT:  Chappell, Joseph
;    APPLICANT:  Yin, Shaohui
;    APPLICANT:  Cornett, Catherine A.G.
;    TITLE OF INVENTION:  Transcriptional Control Sequences and
;    TITLE OF INVENTION:  Methods
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Greenlee and Winner, P.C.
;      STREET:  5370 Manhattan Circle, Suite 201
;      CITY:  Boulder
;      STATE:  CO
;      COUNTRY:  US
;      ZIP:  80303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/443,639
;      FILING DATE:  18-MAY-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ferber, Donna M.
;      REGISTRATION NUMBER:  33878
;      REFERENCE/DOCKET NUMBER:  69-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (303) 499-8080
;      TELEFAX:  (303) 499-8089
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4254 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Nicotiana tabacum
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(1217..1327, 1455..1718, 1806..2182,
;      LOCATION:  2259
;      LOCATION:  ..2477, 2609..2747, 2903..3148, 3262..3558)
US-08-443-639-7

  Query Match             3.1%;  Score 41;  DB 2;  Length 4254;
  Best Local Similarity   52.7%;  Pred. No. 0.098;
  Matches   89;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     265 TATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTA 324
           | |||| |||  |||| |||||   | | ||||||    |||| | | || ||  |  ||
Db     367 TGTACACACACACACACATACATTTATATATATATATATATATATATATATATATA 308

Qy     325 AAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTT 384
           |  || | ||   || || | ||  |          ||| |  |    | | ||  |  ||
Db     307 TATATATATATATATATATATATATTAATTCTATAATTTTCTTTAACTTTTCACAATTTT 248

Qy     385 TACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATA 433
           |||| | |    || |||   || |||     | || | || |||
Db     247 AGTGTATTTTTCTGATTTATATAATGATTTTATAAATTTATAAAATATA 199


RESULT 12
US-08-973-462-1/c
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 6152
;   TYPE: DNA
;   ORGANISM: P. falciparum
US-08-973-462-1

  Query Match             3.1%;  Score 40.6;  DB 4;  Length 6152;
  Best Local Similarity   54.3%;  Pred. No. 0.16;
  Matches   82;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     199 TAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACG 258
           || | || ||| |  |||||| || | |  | || | |||   | | ||     ||||
Db    6009 TAAGCTAAATTTTTATCTTCTAATATATATACATATATATATATATATATTATATAAT 5950

Qy     259 AGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATT 318
           |  | || ||   | |  || |||  ||  || ||||||||||  |  | |  | || | |
Db    5949 ATAAATTTTAATAAAACACAAATAATTAATATTACATATATCGTATGAATAATATACAGT 5890

Qy     319 AAATTAAAACTAAAAATGACTAAATTTCTAA 349
           ||| |   |   ||   | ||||  ||
Db    5889 TTCTTACAGAAATCTATATATTAATTAAAAA 5859


RESULT 13
US-08-599-252-96
; Sequence 96, Application US/08599252
; Patent No. 5705343
;  GENERAL INFORMATION:
;    APPLICANT:  DRAYNA, DENNIS T.
;    APPLICANT:  FEDER, JOHN N.
;    APPLICANT:  GNIRKE, ANDREAS
;    APPLICANT:  KIMMEL, BRUCE E.
;    APPLICANT:  THOMAS, WINSTON J.
;    APPLICANT:  WOLFF, ROGER K.
;    TITLE OF INVENTION:  METHOD TO DIAGNOSE HEREDITARY
;    TITLE OF INVENTION:  HEMOCHROMATOSIS
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
```

```
;       STREET:  2000 Pennsylvania Ave. N.W., Suite 5500
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20006-1888
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/599,252
;       FILING DATE:  09-FEB-1996
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MURASHIGE, KATE H.
;       REGISTRATION NUMBER:  29,959
;       REFERENCE/DOCKET NUMBER:  9053-0001.21
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 887-1500
;       TELEFAX:  (202) 887-0763
;       TELEX:  90-4030
;   INFORMATION FOR SEQ ID NO:  96:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1000 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-599-252-96

  Query Match             3.1%;  Score 40.4;  DB 1;  Length 1000;
  Best Local Similarity   46.4%;  Pred. No. 0.067;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy   208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
Db   117 TTATTTATTATTTATTAAAAATAATATTTAATATTAAATATTATTTATTACTTATATTAT 176

Qy   268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
Db   177 TTATTATTATAATAATTTATAATATGTCACATTATAAAATATTATTTAATTAAAATTTAA 236

Qy   328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
Db   237 TGTCACTCTATTATTTTAATATCATAAAATACATGAAAATACATTTTTCTGTAGAATCAC 296

Qy   388 CTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAA 447
Db   297 GTTTTCCTCCTACTGTGAATAAAGACATAACTCACTAAGGGGAAGAATCTTGGCCCNAAG 356

Qy   448 ACAGTGCTTTGTACGACTATATA 470
Db   357 TGTGTGATAAATCANANAANANA 379


RESULT 14
PCT-US96-06352-96
; Sequence 96, Application PC/TUS9606352
;  GENERAL INFORMATION:
;     APPLICANT:  DRAYNA, DENNIS T.
;     APPLICANT:  FEDER, JOHN N.
;     APPLICANT:  GNIRKE, ANDREAS
;     APPLICANT:  KIMMEL, BRUCE E.
;     APPLICANT:  THOMAS, WINSTON J.
;     APPLICANT:  WOLFF, ROGER K.
;     TITLE OF INVENTION:  METHOD TO DIAGNOSE HEREDITARY
;     TITLE OF INVENTION:  HEMOCHROMATOSIS
;     NUMBER OF SEQUENCES:  124
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORRISON & FOERSTER
;       STREET:  2000 Pennsylvania Ave. N.W., Suite 5500
;       CITY:  Washington
;       STATE:  DC
;       COUNTRY:  USA
;       ZIP:  20006-1888
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US96/06352
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/599,252
;       FILING DATE:  09-FEB-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MURASHIGE, KATE H.
;       REGISTRATION NUMBER:  29,959
;       REFERENCE/DOCKET NUMBER:  9053-0001.21
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 887-1500
;       TELEFAX:  (202) 887-0763
;       TELEX:  90-4030
;   INFORMATION FOR SEQ ID NO:  96:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1000 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
PCT-US96-06352-96

  Query Match             3.1%;  Score 40.4;  DB 5;  Length 1000;
  Best Local Similarity   46.4%;  Pred. No. 0.067;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy   208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
Db   117 TTATTTATTATTTATTAAAAATAATATTTAATATTAAATATTATTTATTACTTATATTAT 176

Qy   268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
Db   177 TTATTATTATAATAATTTATAATATGTCACATTATAAAATATTATTTAATTAAAATTTAA 236

Qy   328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
Db   237 TGTCACTCTATTATTTTAATATCATAAAATACATGAAAATACATTTTTCTGTAGAATCAC 296

Qy   388 CTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAA 447
Db   297 GTTTTCCTCCTACTGTGAATAAAGACATAACTCACTAAGGGGAAGAATCTTGGCCCNAAG 356

Qy   448 ACAGTGCTTTGTACGACTATATA 470
Db   357 TGTGTGATAAATCANANAANANA 379


RESULT 15
PCT-US96-06583-96
; Sequence 96, Application PC/TUS9606583
;  GENERAL INFORMATION:
;     APPLICANT:  DRAYNA, DENNIS T.
;     APPLICANT:  FEDER, JOHN N.
;     APPLICANT:  GNIRKE, ANDREAS
;     APPLICANT:  KIMMEL, BRUCE E.
;     APPLICANT:  THOMAS, WINSTON J.
;     APPLICANT:  WOLFF, ROGER K.
;     TITLE OF INVENTION:  METHOD TO DIAGNOSE HEREDITARY
;     TITLE OF INVENTION:  HEMOCHROMATOSIS
;     NUMBER OF SEQUENCES:  124
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MORRISON & FOERSTER
;       STREET:  2000 Pennsylvania Ave. N.W., Suite 5500
;       CITY:  Washington
;       STATE:  DC
```

```
;       COUNTRY:  USA
;       ZIP:  20006-1888
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US96/06583
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/599,252
;       FILING DATE:  09-FEB-1996
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  MURASHIGE, KATE H.
;       REGISTRATION NUMBER:  29,959
;       REFERENCE/DOCKET NUMBER:  9053-0001.21
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 887-1500
;       TELEFAX:  (202) 887-0763
;       TELEX:  90-4030
;   INFORMATION FOR SEQ ID NO:  96:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1000 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
PCT-US96-06583-96

  Query Match             3.1%;  Score 40.4;  DB 5;  Length 1000;
  Best Local Similarity   46.4%;  Pred. No. 0.067;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy    208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
          || | |  |  |||||       || |||| | ||   || |  ||||        ||||
Db    117 TTATTTATTATTTATTAAAAATAATATTTAATATTAAATATTATTTATTACTTATATTAT 176

Qy    268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
            ||   |  |   |  ||| ||     | || | || || ||  | |||||| |  ||
Db    177 TTATTATTATAATAATTTATAATATGTCACATTATAAAATATTATTTAATTAAAATTTAA 236

Qy    328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
              |    | ||  ||  || ||  ||  ||||   |||   || | |   |   | ||
Db    237 TGTCACTCTATTATTTTAATATCATAAAATACATGAAAATACATTTTTCTGTAGAATCAC 296

Qy    388 CTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAA 447
            | |||  | |  |   | |    |  | ||||  |||     | | |||       ||
Db    297 GTTTTCCTCCTACTGTGAATAAAGACATAACTCACTAAGGGGAAGAATCTTGGCCCNAAG 356

Qy    448 ACAGTGCTTTGTACGACTATATA 470
              ||| |   |   |  | | |
Db    357 TGTGTGATAAATCANANAANANA 379


Search completed: November  7, 2002, 10:47:55
Job time : 105.015 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 08:30:58 ; Search time 1237.97 Seconds
(without alignments)
14282.232 Million cell updates/sec

Title: US-09-905-558C-3
Perfect score: 1310
Sequence: 1 cccatcgctgctttgtctac.....caagggaaagtgatcggatg 1310

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	149.2	11.4	374	12	BH255115	BH255115 RXL1BAM00
c	2	123.8	9.5	223	12	AZ921021	AZ921021 1006023E0
c	3	107.6	8.2	496	9	AW928053	AW928053 945007D02
	4	58.2	4.4	572	10	BI885732	BI885732 fn12e11.y
c	5	58.2	4.4	1101	12	CNS0106X	AL098595 Drosophil
c	6	57.8	4.4	500	9	AU087781	AU087781 AU087781
c	7	57.4	4.4	935	12	CNS006XK	AL066051 Drosophil
	8	56.6	4.3	534	12	BH255739	BH255739 LDH5BAM00
	9	55.6	4.2	1058	12	CNS014D0	AL103998 Drosophil
c	10	55.2	4.2	818	10	BI952390	BI952390 HVSMEm000
	11	55	4.2	925	12	CNS0091P	AL053013 Drosophil
	12	54.8	4.2	936	10	BG852371	BG852371 1024034A0
	13	54	4.1	798	12	AQ251130	AQ251130 F26H6-Sp6
	14	54	4.1	1101	12	CNS00FVE	AL071298 Drosophil
c	15	53.8	4.1	481	12	CNS016IP	AL106795 Drosophil
c	16	53.8	4.1	848	10	BI950718	BI950718 HVSMEl002
	17	53.8	4.1	1038	12	CNS015VV	AL105973 Drosophil
c	18	53.4	4.1	802	12	CNS02R43	AL210108 Tetraodon
	19	53.4	4.1	1101	12	CNS008WC	AL052719 Drosophil
	20	53.2	4.1	945	12	CNS04D0K	AL285149 Tetraodon
c	21	53.2	4.1	1082	10	BE455154	BE455154 HVSMEh009
c	22	53	4.0	914	12	AZ674804	AZ674804 ENTKN64TR
c	23	53	4.0	928	12	CNS00DKY	AL071865 Drosophil
	24	53	4.0	1101	12	CNS0039G	AL063921 Drosophil
	25	52.8	4.0	652	12	BH255629	BH255629 LDH5BAM00
	26	52.8	4.0	928	12	CNS00DKY	AL071865 Drosophil
c	27	52.6	4.0	1092	12	CNS020K7	AL175696 Tetraodon
c	28	52.6	4.0	1101	12	CNS017KE	AL108152 Drosophil
c	29	52.6	4.0	1101	12	CNS017ZQ	AL108704 Drosophil
c	30	52.4	4.0	479	10	BM163977	BM163977 EST566500
c	31	52.4	4.0	745	12	AZ046300	AZ046300 nbeb0091L
	32	52.2	4.0	914	12	CNS002JY	AL097768 Drosophil
c	33	52.2	4.0	1101	12	CNS00H18	AL072804 Drosophil
c	34	52	4.0	861	10	BI953364	BI953364 HVSMEm001
	35	52	4.0	1101	12	CNS017EJ	AL107941 Drosophil
	36	51.8	4.0	649	12	B19263	B19263 T22C5-T7 TA
c	37	51.8	4.0	1043	12	CNS0145P	AL103735 Drosophil
c	38	51.8	4.0	1101	12	CNS00Z3I	AL097176 Drosophil
	39	51.4	3.9	553	9	AI657313	AI657313 486093C05
c	40	51.4	3.9	563	10	BG004199	BG004199 QV4-GN012
c	41	51.4	3.9	776	12	AG130889	AG130889 Pan trogl
c	42	51.4	3.9	910	12	CNS006ON	AL065629 Drosophil
	43	51.4	3.9	925	10	BG441241	BG441241 GA__Ea001
	44	51.4	3.9	1101	12	CNS016LI	AL106896 Drosophil
c	45	51.2	3.9	342	10	BM168596	BM168596 EST571119

ALIGNMENTS

RESULT 1
BH255115/c
LOCUS BH255115 374 bp DNA linear GSS 29-NOV-2001
DEFINITION RXL1BAM0001A01r Zea mays L. XL1 methyl filtration maize root genomic shotgun library Zea mays genomic clone RXL1BAM0001A01r, DNA sequence.
ACCESSION BH255115
VERSION BH255115.1 GI:17150008
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 374)
AUTHORS Kim,S.W., Yu,Y., Lee,M.C., Yang,T.J., Main,D., Henry,D., Oates,R. and Wing,R.A.
TITLE Genomic shotgun library from maize
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACCCTCACTAAAGGG
Class: shotgun
High quality sequence stop: 374.
FEATURES Location/Qualifiers
source 1. .374
/organism="Zea mays"
/strain="B73"
/db_xref="taxon:4577"
/clone="RXL1BAM0001A01r"
/clone_lib="Zea mays L. XL1 methyl filtration maize root genomic shotgun library"
/tissue_type="Root"
/lab_host="XL1Blue"
/note="Vector: pCUGIblu-1; Site_1: Sau3A1;

Methyl-filtration library, Nuclei DNA was digested with Sau3A1, size fractionated and transformed to E.Coli.XL1Blue."
BASE COUNT 125 a 64 c 70 g 115 t
ORIGIN

Query Match 11.4%; Score 149.2; DB 12; Length 374;
Best Local Similarity 79.9%; Pred. No. 4.9e-21;
Matches 214; Conservative 0; Mismatches 48; Indels 6; Gaps 3;

```
Qy      78 CCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGTACGGCTACATCGAA 137
           || |||||| |||| ||  |||||||| |  ||||||||||||   ||   ||| ||||||
Db     356 CCATTCGAGGGACTACACTGCGTACATTTGCCTGCATCGACTGCATA---CTAAATCGAA 300

Qy     138 CATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCC 197
           |||| |||||||||||||| |||||||||||| ||   ||| ||| |||||||||  || |
Db     299 CATACACACGAGATGTCTTGTGTGAATAGAGCCAACGATGTCTTGAGCATCGGTCCCTTC 240

Qy     198 GTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATAC 257
           | |||||||||| |||||||  |  ||||||||  |||||  |||||||||| | |||| |
Db     239 GCAGGGTACATTATGTTCTTAGTGATTGTGCAT-GTTTTACTGTTGTTTACCGCTTATGC 181

Qy     258 GAGTAGTTATACATACATGCAC--ATACATATCATCACATATATCACAATATTTTTCTAA 315
           |||||||  |||| ||||||||  |||||| ||||||||||| ||||||  |  ||| ||
Db     180 GAGTAGTAGTACACACATGCACATATACATGTCATCACATACATCACACTGGTTTCCTGG 121

Qy     316 ATTAAATTAAAACTAAAAATGACTAAAT 343
           |||||||||||||| | ||||||||  ||
Db     120 ATTAAATTAAAACTGATAATGACTATAT 93
```

RESULT 2
AZ921021/c
LOCUS AZ921021 223 bp DNA linear GSS 17-DEC-2001
DEFINITION 1006023E04.2EL_y1 1006 - RescueMu Grid G Zea mays genomic, DNA sequence.
ACCESSION AZ921021
VERSION AZ921021.1 GI:13392238
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 223)
AUTHORS Walbot,V.
TITLE Maize genomic sequences found using engineered RescueMu transposon
JOURNAL Unpublished (2001)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1006023 row: 39
Class: transposon-tagged.
FEATURES Location/Qualifiers
source 1. .223
/organism="Zea mays"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/clone_lib="1006 - RescueMu Grid G"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: leaf; Vector: RescueMu (engineered from pBlueScript backbone); Site_1: BamHI; Site_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."
BASE COUNT 73 a 36 c 46 g 68 t
ORIGIN

Query Match 9.5%; Score 123.8; DB 12; Length 223;
Best Local Similarity 82.4%; Pred. No. 9.5e-16;
Matches 154; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

```
Qy     166 GAGTCACTAATGCCTTAAGCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTG 225
           ||  ||||||||| || |||||||||   ||||||||||||||||||||||||  || |||
Db     223 GATCCACTAATGCTTTGAGCATCGGTCCATCCGTAGGGTACATTCTGTTCTTAGTAATTG 164

Qy     226 TGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACAT 285
           |||||  ||||| || ||||||||| |||| |||||| | ||||| | ||||||||| |||
Db     163 TGCAT-GTTTTAATGCTGTTTACTGCTTATGCGAGTATTCATACACATATGCACATGCAT 105

Qy     286 ATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTT 345
             | ||||||| |||||| || ||||||  |||||||||||||| |||||||  |||| | |
Db     104 GCCGTCACATACATCACACTAATTTTCTGGATTAAATTAAAACAAAAAATGCCTAATTCT 45

Qy     346 CTAACAC 352
           |||||||
Db      44 CTAACAC 38
```

RESULT 3
AW928053/c
LOCUS AW928053 496 bp mRNA linear EST 30-MAY-2000
DEFINITION 945007D02.Y1 945 - Mixed adult tissues from Walbot lab, same as 707 (SK) Zea mays cDNA, mRNA sequence.
ACCESSION AW928053
VERSION AW928053.1 GI:8103414
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 496)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 945007 row: D column: 02.
FEATURES Location/Qualifiers
source 1. .496
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="945 - Mixed adult tissues from Walbot lab, same as 707 (SK)"
/tissue_type="tassel, kernal, silk, husk, root, leaf"
/dev_stage="fully-grown"
/lab_host="DH10B"
/note="Organ: tassel, kernal, silk, husk, root, leaf; Vector: pGAD10; Site_1: EcoRI; cDNA library from fully differentiated maize tissues from an active Mutator plant. Tissue ratio is 4:2:1:1:1:1 (tassel, kernel, silk, husk, root, leaf). Unidirectionally cloned. New library number given to library 707 for additional sequencing."

```
BASE COUNT      203 a     62 c    107 g    124 t
ORIGIN

  Query Match             8.2%;  Score 107.6;  DB 9;  Length 496;
  Best Local Similarity   78.0%;  Pred. No. 2.4e-12;
  Matches  142;  Conservative    0;  Mismatches   39;  Indels    1;  Gaps    1;

Qy     208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
           || | | |||| |||||||||||  ||| |||| |||||||||| |||| ||||||||||
Db     429 TTTTTTCCTTCGTATTTGTGCAT-GTTTCATTGCTGTTTACTGCTTATGCGAGTAGTTAT 371

Qy     268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
           || |||||||||||||| |||||||||||||| || | |||||||  ||||||||||||
Db     370 GCACACATGCACATACATGTCATCACATATATCGCACTGTTTTTCTGGATTAAATTAAAA 311

Qy     328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
           ||||||||| |||| |||||||||   || | | | |  ||  | |    ||| |  |||
Db     310 CTAAAAATGCCTAACTTTCTAACAAATACTAAAGTTTGTTGCGTCCAAACACATCAATAC 251

Qy     388 CT 389
           ||
Db     250 AT 249


RESULT 4
BI885732
LOCUS       BI885732                572 bp    mRNA    linear   EST 12-OCT-2001
DEFINITION  fn12e11.y1 Zebrafish WashU MPIMG EST Danio rerio cDNA clone 4490204
            5' similar to contains element MSR1 repetitive element ;, mRNA
            sequence.
ACCESSION   BI885732
VERSION     BI885732.1  GI:16093003
KEYWORDS    EST.
SOURCE      zebrafish.
  ORGANISM  Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
            ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
            ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
            Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
            Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
            and Wilson,R.
  TITLE     WashU Zebrafish EST Project 1998
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Stephen L. Johnson
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: zbrafish@watson.wustl.edu
            cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
            Matthew Clark. DNA Sequencing by: Washington University Genome
            Sequencing Center Clone distribution: Genome Systems, St. Louis,
            Missouri (web address: www.genomesystems.com) (email contact:
            info@genomesystems.com) and Research Genetics, Huntsville, Alabama
            (web address: www.resgen.com) (email contact: info@resgen.com) and
            RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
            www.rzpd.de)
            Seq primer: T3 ET from Amersham
            High quality sequence stop: 366.
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Danio rerio"
                     /db_xref="taxon:7955"
                     /clone="4490204"
                     /clone_lib="Zebrafish WashU MPIMG EST"
                     /sex="mixed"
                     /tissue_type="26 somite embryos, adult livers, shield
                     stage embryos"
                     /lab_host="XL1-blue MRF"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
                     strand cDNA was primed with a Not I - oligo(dT)15 primer
                     [5'pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT3']
                     double-stranded cDNA was ligated to Sal I adaptors (BRL),
                     digested with Not I and cloned into the Not I and Sal I
                     sites of the pSPORT1 vector (BRL). Library was constructed
                     by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
                     Institut fuer Molekulare Genetik,Berlin). cDNAs for EST
                     analysis were selected following oligonucleotide
                     hybridization fingerprinting of arrayed clones from
                     zebrafish late somitogenesis (26 ss), adult liver or
                     embryonic shield stage (5.6 h) libraries.  Fingerprint
                     data were used to computationally cluster cDNAs, and a
                     single cDNA from each cluster was chosen for sequencing.
                     In some cases multiple members of the same cluster were
                     sequenced to assess clustering parameters or single clones
                     were sequenced additional times to assess quality
                     control."
BASE COUNT      13 a    421 c     44 g     16 t     78 others
ORIGIN

  Query Match             4.4%;  Score 58.2;  DB 10;  Length 572;
  Best Local Similarity   43.6%;  Pred. No. 0.048;
  Matches  159;  Conservative    0;  Mismatches  206;  Indels    0;  Gaps    0;

Qy     702 CCGCCTTTATTTCAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACC 761
           || ||           |   |   ||||  ||| | |  |   || |      | |  ||
Db      76 CCCCCCNNNNNNNNNNCNNNCCCCCCCCCGGGGGGGGGGGGCCGCNCNCNCCCCCCCCCC 135

Qy     762 CCAACCTCTCACCGAAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGC 821
           ||  || | | ||       || |||  |   | |    |   |          ||||   |
Db     136 CCNCCCCCCCCCCNCCCCNCCCCGCGCGGCGGCCCCCCCCCCCCCCNNCCCCCCCCCCCCC 195

Qy     822 CCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGC 881
           || | || |  || |  | ||  ||| |  |  || | | ||  ||   ||    | |
Db     196 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCNCCCCNCCCC 255

Qy     882 GGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCGTGTCG 941
             | | |  || |||||  |||   | |    ||||  |||       |||    |
Db     256 CCCCCNCCCCCCCCCCCCCCNCCCCCNCCCCCCCCCCCCCCCCCCCCNNNCCCCCCCCCCC 315

Qy     942 GTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTG 1001
             || | | |  |  |    |  ||| | |  |     ||||    | |||    |||
Db     316 CCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCNNNCCCCCC 375

Qy    1002 TCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCACGGCCGACTGCGCTGCGC 1061
           |  |||| ||||  ||  | |      | |  |  ||||| | ||   || | |
Db     376 CCCCCCCTCTGTTGTTTGTTTTCCNCCCCCCCNCCCCCCCCCCCGCTTTCTTCCCGCGCN 435

Qy    1062 CGCCC 1066
           | |||
Db     436 CCCCC 440


RESULT 5
CNS0106X/c
LOCUS       CNS0106X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098595
VERSION     AL098595.1  GI:5610206
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL'  Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
```

```
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03K20"
                     /note="end : T7"
BASE COUNT      258 a    107 c     60 g    175 t    501 others
ORIGIN

  Query Match             4.4%;  Score 58.2;  DB 12;  Length 1101;
  Best Local Similarity   19.2%;  Pred. No. 0.049;
  Matches   87;  Conservative  182;  Mismatches  184;  Indels    0;  Gaps    0;

Qy    209 TCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATA 268
          |:| ||::|:|   |  :  :|:|:|||::: |   |  :|  :|:|:    |: :|: :
Db   1027 TMTMTTHHTMTMTMTMTHTTMTMTHTTTHHHTTMTMTMTMTMTHTHTMTMTWTMTHTHMM 968

Qy    269 CATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAAC 328
          :::::::| :::::::::|::: ::::: : :|: ::::: ::: ::::::|:: :|  : ::
Db    967 HMHMHMTMHMMHHHMMTMHMTYHMMWMMMHTMTHMMHHMHHHMHHMHMTHMTMTMTMTHM 908

Qy    329 TAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACC 388
          ::::::   :|::: |  |::: :   :  ::: | |::|  |:|:|:::::|  :: |: :
Db    907 HMMMMMMMTMTMHMMTMATMMTMATTMTHHHHMTMTMMTTMTHTMTMHMHCTYYYMTMTH 848

Qy    389 TATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAA 448
            |:   : || : :|:  :|  ::  ::: :::| ::: :|  : |:| : :|  :  :
Db    847 CTTHTMTMTTTMHMMTMMMHCWTMHMMMMMMYHMTYMMMCHCYCHHTHTTHTMTMTHTTM 788

Qy    449 CAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAGTCGTT 508
           : |  :|: | :  :|:|   : :  : :     : : : :|::  |::| |: |   :
Db    787 THMTHTHTHHTTMTTMTMTYMYMHMMTHWYYMHYTHTMTMTMTMHMTCMHTTTHTTATAW 728

Qy    509 TGAAGATTGAACCTATATATCGGTACGGTTAATCCGTCTATGTACGTGGGCATGACGAAC 568
            |:  :|   : |:|:|:|:  : :   ||:|  |:|:|  ::      :: :  ::
Db    727 ATAHWTYTMTYMATMTMTMTMTMYTHMCATAMTTTMTMTMTTMMMTMTMATMHWMWTHMT 668

Qy    569 ACCCGTGATAACGAAGGATTAACGTGCACAATCATAAATCCAAAGTAGGAGCGGTGCATG 628
          |::: : : :|   |   |  |:    |::  :|| |: :  |:  | :| :  | :|
Db    667 AHMHMYTMMMATMTATATTAYAHMAHTAYMTAMATTAWAYATAMACATKATMAMTMHAAC 608

Qy    629 ATGAGAATCGCTCTCAGTACTCGACATAATGAA 661
          |     :: :   : :  |:: || |:::  |:
Db    607 AMACMTMWAKAAAWAMAAAMDAGAGADMMACAR 575


RESULT 6
AU087781/c
LOCUS       AU087781                500 bp    mRNA    linear   EST 27-JAN-2001
DEFINITION  AU087781 Sugano Malaria cDNA library Plasmodium falciparum cDNA
            clone XPFn5810, mRNA sequence.
ACCESSION   AU087781
VERSION     AU087781.1  GI:12389922
KEYWORDS    EST.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
  TITLE     FULL-malaria: a database for a full-length enriched cDNA library
            from human malaria parasite, Plasmodium falciparum
  JOURNAL   Nucleic Acids Res. 29 (1), 70-71 (2001)
  MEDLINE   20574754
COMMENT     Contact: Junichi Watanabe
            Institute of Medical Science
            The University of Tokyo, Department of Parasitology
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5378
            Fax: 81-3-5449-5410
            Email: jwatanab@manage.ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /clone="XPFn5810"
                     /clone_lib="Sugano Malaria cDNA library"
                     /dev_stage="erythrocytic stage"
BASE COUNT      228 a     25 c     33 g    203 t     11 others
ORIGIN

  Query Match             4.4%;  Score 57.8;  DB 9;  Length 500;
  Best Local Similarity   54.2%;  Pred. No. 0.058;
  Matches  136;  Conservative    0;  Mismatches  114;  Indels    1;  Gaps    1;

Qy    220 TATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 279
          ||| | |  |||| |||| |  | |||| | |||  |   ||| ||| | ||| || |||
Db    274 TATATATATATATATTTACTAATTTTTATTTATTTAAAATGTAATTAAAAATATATACAC 215

Qy    280 ATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACT 339
          ||| ||||    | ||||||    |||| | | || ||  | |||    ||    |   |
Db    214 ATATATATATATATATATATATATATATATATATATATATTTATGTGTATTTTTTTTT 155

Qy    340 AAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATT 399
           | | | |||    ||    | | |||||| | | | |  | | ||  ||||  | |
Db    154 TATTAT-TAAATTTAATTTTAATCTAATGTGTGTTTCTAATAATATAAATATTANATAAA 96

Qy    400 GTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGT 459
           || | |||  ||  |    | | |   || | | |||   |  |||| | | ||   |
Db     95 TTTTTTTTTATAAAATNCGACGAAAGTGAAAAATGACTATTGTAAAAAATATTTCTGGAT 36

Qy    460 ACGACTATATA 470
          ||    ||| |
Db     35 ACATAAATACA 25


RESULT 7
CNS006XK/c
LOCUS       CNS006XK                935 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC #
            BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL066051
VERSION     AL066051.1  GI:4945019
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 935)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
```

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .935
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR14N09"
                     /note="end : T7"
BASE COUNT      257 a    170 c    162 g     96 t    250 others
ORIGIN
```

```
  Query Match             4.4%;  Score 57.4;  DB 12;  Length 935;
  Best Local Similarity   30.2%;  Pred. No. 0.072;
  Matches  115;  Conservative   86;  Mismatches  180;  Indels    0;  Gaps    0;

Qy     804 CGAGGTGGTGCCCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCA 863
           :| :| |:   : :| :||| : : || |:  | :: : ::  ||| : :|  :||| :
Db     921 SGSSGCGSGCGSGSGGSCCCCSGSCCCSCSCGCCSSCSCSSSSCCCSSCSCGCSCGCCSC 862

Qy     864 CCAGCCATGCCGTGTCGCGGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTC 923
           :|| ||   | ::  |  |:|  |  || :|  :  ||: : | ||   :|:||
Db     861 GGSSCCSCGCGCSGSSSGCGCCCSGGSCGCCCGSCGGSCSCCSSSCCGCGCCGSCSCGCG 802

Qy     924 CCGGCCGTGCCCGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCC 983
           |||||:|  |: | | :|   |: ||  ::|  |: ::  || ||:|   : : |  :|
Db     801 CCGGCSGGCCSGGCGSSGCGGCSGGCCGSSCCGCSGKSKCGCGCGSGGSCSSSSGSGSCG 742

Qy     984 TGGACACCCGACACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCA 1043
           :| : |::  : |:|   ::|: |   |      |   |:                : :| |
Db     741 GSGGSGCSSCSSGCSCGCGSSGSGCGSCGCCCSCGCSCGCSCGCSSGCGSCGCSGSCGCG 682

Qy    1044 CGGCCGACTGCGCTGCGCCGCCCGGATATATATACCCATCGTTATCGATCGATCGATCGC 1103
           | |||:    |:|  |:|::::||:         |::  ::    ||  |:  ||  |:|
Db     681 CCGCCSCGCCCSCCCCSCSSSSCGSSCSSCCSSSCSSCCSSSGCCCGCCCSCGCGCGCSC 622

Qy    1104 GTCACTCACGGGTAGCTCATGGTCGAGCGTAGCATGCAGGAACTTATTTGCCGTGCGCTC 1163
           | | | | | || ||| || |  :|:  :  : | :::::  :       : || |||: |
Db     621 GCCGCSCCCCGGCAGCSCASGCSSGMAGVAGSGASRSVVSGGSSSSGSASGCGCGCGMGC 562

Qy    1164 CCAGGTCTCCGCTCGCGTGCC 1184
              :|:: : |:  :||  |:
Db     561 RAGSGKMGSAGSSGRCGGACS 541
```

```
RESULT 8
BH255739
LOCUS       BH255739                534 bp    DNA     linear   GSS 29-NOV-2001
DEFINITION  LDH5BAM0003C03f Zea mays L. DH5alpha methy filtration maize leaf
            genomic shotgun library Zea mays genomic clone LDH5BAM0003C03f, DNA
            sequence.
ACCESSION   BH255739
VERSION     BH255739.1  GI:17150632
KEYWORDS    GSS.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 534)
  AUTHORS   Kim,S.W., Yu,Y., Lee,M.C., Yang,T.J., Main,D., Henry,D., Oates,R.
            and Wing,R.A.
  TITLE     Genomic shotgun library from maize
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: shotgun
            High quality sequence stop: 531.
FEATURES             Location/Qualifiers
     source          1. .534
                     /organism="Zea mays"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="LDH5BAM0003C03f"
                     /clone_lib="Zea mays L. DH5alpha methy filtration maize
                     leaf genomic shotgun library"
                     /tissue_type="Leaf"
                     /lab_host="DH5alpha"
                     /note="Vector: pCUGIblu-1; Site_1: Sau3A1; Site_2: Sau3A1;
                     Methyl-filtration library, Nuclei DNA was digested with
                     Sau3A1, size  fractionated and transformed to
                     E.Coli.DH5alpha."
BASE COUNT      174 a     77 c     78 g    205 t
ORIGIN
```

```
  Query Match             4.3%;  Score 56.6;  DB 12;  Length 534;
  Best Local Similarity   58.2%;  Pred. No. 0.1;
  Matches  139;  Conservative    0;  Mismatches   94;  Indels    6;  Gaps    2;

Qy     204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
           | || | | |||| ||     || | ||| | |  |   |||||  |  |  |  ||||||
Db     236 TTCAGTTTTTTCTACTATCATGAGAATAATATATTGTATGTTTGTTACTATTTAGAGTAG 295

Qy     264 TTATACATACATGCACATACATATCATCACATATATCACAATAT-TTTTCTAAATTAAAT 322
           |      | |||||||| || || ||||||||||| | |    || |||| |  |||||||
Db     296 T-----ACACATGCACGTAAATGTCATCACATATTTTATGCAATGTTTTTTGGATTAAAT 350

Qy     323 TAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAAC 382
           |||     ||  || |||||||||||||||||| ||||   | | ||        |  ||
Db     351 TAAGCAATAATTTGCCTAAATTTCTAACACCAACTACATAAGATTTTTCATGTTTAATAC 410

Qy     383 TTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAA 441
           ||  |  |  |||  |  ||         |||  | | ||||   || || |   | ||
Db     411 TTGTCAAAAGCTAGTTAATTAGGAGGGTTATTATAGTGTATAGGAAATATGGAGAATAA 469
```

```
RESULT 9
CNS014D0
LOCUS       CNS014D0                1058 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN1iI17 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103998
VERSION     AL103998.1  GI:5615609
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1058)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
```

```
          pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1058
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN11I17"
                     /note="end : SP6"
BASE COUNT      121 a    175 c     93 g    396 t    273 others
ORIGIN

  Query Match             4.2%;  Score 55.6;  DB 12;  Length 1058;
  Best Local Similarity   32.9%;  Pred. No. 0.17;
  Matches   78;  Conservative   65;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     186 ATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTT 245
           || |: :| || :   : ||:: |:|::| :|::    |:   |: : | |:||  | |
Db     700 ATAGSCWATTCWSASTRATASWMTSTSWTWBTBWVKRSTRATKAKTBATCTVTTWVTATA 759

Qy     246 TACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACAAT 305
           : :|:::: ||| :  : | :|  |    : :  |  ||:   ||| ::|||||:    :
Db     760 HSMTSWWYTTACCMTSMCTSWTSTACCACWCYCTACCCAWTATATCTMMTATATHCACTW 819

Qy     306 ATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTA 365
           |  || |: :|:|| |:|  |  :|| ||: | |  |: |:   : :|  : ::|   :|
Db     820 ACATTSCYTMAHTAYAHTTCACTWAANAAYTATTSTAWATYMTTSTSCTTSWVSAAAAYA 879

Qy     366 ATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTA 422
             ::||:  |::  |    :|::|:| :|| ||: |  ::::|: | |:|:|:  ||
Db     880 TASWTTBTACSMTTACTSAYASYTHTAMTATATYATAAWWHYTMWATTHTYAHCATA 936


RESULT 10
BI952390/c
LOCUS       BI952390                818 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  HVSMEm0006C23f Hordeum vulgare green seedling EST library
            HVcDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone
            HVSMEm0006C23f, mRNA sequence.
ACCESSION   BI952390
VERSION     BI952390.1  GI:16296163
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
            Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons
            ,J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected Morex (compatible) seedling
            cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 224
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 808.
FEATURES             Location/Qualifiers
     source          1. .818
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEm0006C23f"
                     /clone_lib="Hordeum vulgare green seedling EST library
                     HVcDNA0014 (Blumeria infected)"
                     /tissue_type="green seedling leaf"
                     /lab_host="TJC121"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; Morex (mla) plants were greenhouse grown in the R
                     Wise lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were infected with isolate 5874 of
                     Blumeria graminis f. sp. hordei, and leaves were harvested
                     24, 48 and 72 hr post-inoculation and snap frozen (Wise).
                     In the TJ Close lab at the University of California,
                     Riverside, total RNA was prepared from each sample pool,
                     equal quantities of all three RNA pools were combined,
                     poly(A) RNA was purified from the mixture, one primary
                     unamplified cDNA library was made, and 1 million pfu were
                     in vivo excised to give pBluescript SK(-) cDNA phagemids
                     (Chin). Phagemids were plated and picked at the Clemson
                     University Genomics Institute (CUGI) (Begum, Palmer,
                     Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
                     sequencing and sequence analysis were performed at CUGI
                     (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
                     The sequence has been trimmed to remove vector sequence
                     and contains a minimum of 100 bases of phred value 20 or
                     above. For more details on library preparation and
                     sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT       74 a    142 c    556 g     46 t
ORIGIN

  Query Match             4.2%;  Score 55.2;  DB 10;  Length 818;
  Best Local Similarity   46.4%;  Pred. No. 0.21;
  Matches  180;  Conservative    0;  Mismatches  208;  Indels    0;  Gaps    0;

Qy     702 CCGCCTTTATTTCAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACC 761
           || |||      |   | | |   |||| |   |   |  |  || |   |  | |  ||
Db     625 CCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCTCCTCGCCCCCCCCC 566

Qy     762 CCAACCTCTCACCGAAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGC 821
           ||  |||||| ||       ||   |     |  |     |   |     |     |||   |
Db     565 CCCCCCTCTCCCCCCCCCTCCCTCCCCCCCCCCCCCCGCCCTCTCTTCCCTCCCCCCC 506

Qy     822 CCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGC 881
           || | | || |  || | || ||  ||| |   |  || |   ||  ||  ||    | |
Db     505 CCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCACCCCCCCTCTCCCCCCCCC 446

Qy     882 GGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCGTGTCG 941
             |  | |  || |||||  ||| | | |    ||||  |||  ||  |||    |
Db     445 CCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCT 386

Qy     942 GTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTG 1001
             |||| | |    |||   | |||| | |  |  | |||||    | |||  |||||
Db     385 TCCCGCCCCCCCTCACGCCCCCGCCCCCCCCCCCCCACTCCCCCCCCCCCCCCCCCACCCCC 326

Qy    1002 TCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCACGGCCGACTGCGCTGCGC 1061
           |  |||       ||||  |||   | |   ||  ||| | | ||| ||    | || |
Db     325 CCCCCCCCCCCCGCATTCTCCCCCCCAGCCACTCGCCCCGCTCCGCCCACGCGCCCGCCC 266

Qy    1062 CGCCCGGATATATATACCCATCGTTATC 1089
           |  ||| | |       || ||   |||
Db     265 CCGCCGCACAGCCCCCGCCTTCTCCATC 238


RESULT 11
CNS0091P
LOCUS       CNS0091P                925 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
```

```
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .925
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR19D16"
                      /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN

  Query Match             4.2%;  Score 55;  DB 12;  Length 925;
  Best Local Similarity   13.0%;  Pred. No. 0.23;
  Matches   41;  Conservative  161;  Mismatches  111;  Indels    2;  Gaps    1;

Qy     692 GGACCAGGGGCCGCCTTTATTTCAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAA 751
           :: || :::::::|: |     |  :::::|  :|::: : :::: :  :::|::  :
Db     597 SSYCCSSSBSSSKCSSTSBSCSCCCSSKSVCGTSCSSSSSCSSSSSSTSSSSTSSSTSKS 656

Qy     752 GGGCACTACCCCAACCTCTCACCGAAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGG 811
           :::  :  ::::   :::|:|: :  : |  |:|:||:  ::::   |:   ::::: ::
Db     657 SSSGSSSSSSSSYTTSKSTSASGSGSWSAGGGSGSTGSTSSSSSSSSTSTSSSSVSSGSK 716

Qy     812 TGCCCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCAT 871
           : :::| :::: : : :: :::::  : : ::  ::::: :  :|| | | :: :::
Db     717 SSTBSSGSBSSSGSSSSSSTSSBBSCTSTSSSS--SSSYSSSTCSCCTCCCSYSYSSSTS 774

Qy     872 GCCGTGTCGCGGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGT 931
           ::::|:: : :|  ::::    :::: : |  ||| :   |:   ::|:   :|||:::
Db     775 SSSSTSWGSTSGSSSSSVGTSSSSDSTSTCCSCCCYMCTCCSTYBMBCYTSTSCGGSSSS 834

Qy     932 GCCCGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACC 991
           :   |  :||   ||:::| | :  ::| ::::::|:::: :: : :  ::: :: : ::
Db     835 SGKGGVTKCGCGGCGSSSTNGMBGTSSACSSSSSSCSSSSVSSSSKSSASSSSVSSSGSS 894

Qy     992 CGACACCCTGTCGGC 1006
           : : ::: ::::::
Db     895 GVSSNSSSASKSSSS 909


RESULT 12
BG852371
LOCUS       BG852371                936 bp    mRNA    linear   EST 29-MAY-2001
DEFINITION  1024034A05.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II
            Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION   BG852371
VERSION     BG852371.1  GI:14233555
KEYWORDS    EST.
SOURCE      Chlamydomonas reinhardtii.
  ORGANISM  Chlamydomonas reinhardtii
            Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
            Chlamydomonadaceae; Chlamydomonas.
REFERENCE   1  (bases 1 to 936)
  AUTHORS   Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,
            McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
  TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
            Unicellular System for Analyzing Gene Function and Regulation in
            Vascular Plants; project phase 2
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
FEATURES              Location/Qualifiers
     source           1. .936
                      /organism="Chlamydomonas reinhardtii"
                      /strain="CC-1690 wild type mt+ 21gr"
                      /db_xref="taxon:3055"
                      /clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap
                      II"
                      /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
                      XhoI; This library, constructed by John Davies and Jeffrey
                      McDermott, combines cDNAs from CC-1690 cells grown to
                      mid-log phase in TAP (acetate-containing) medium in the
                      light, TAP medium in the dark, HS (minimal) medium in
                      ambient levels of CO2 and HS medium bubbled with 5% CO2.
                      PolyA mRNA was purified from each sample, pooled and cDNA
                      synthesized. The cDNA was directionally cloned into lambda
                      ZAP II (Stratagene) in the EcoRI (5') and XhoI (3') sites.
                      pBluescript II SK- plasmids were excised from the lambda
                      ZAP clones by superinfection with ExAssist (Stratagene)
                      phage. The library was normalized using method 4 described
                      in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      19 a    537 c    308 g     37 t     35 others
ORIGIN

  Query Match             4.2%;  Score 54.8;  DB 10;  Length 936;
  Best Local Similarity   52.0%;  Pred. No. 0.25;
  Matches  145;  Conservative    0;  Mismatches  133;  Indels    1;  Gaps    1;

Qy     717 GCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACCCCAACCTCTCACCGA 776
           ||  || |||||  || ||| | ||||  |         |      | |  ||   | | |
Db     465 GCGGGCGGGCCCGGCGCGCGGGTTGCGCCCGGCGGCCCCGGCGGCGCGGCCGGCCCCGGC 524

Qy     777 AAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACGTCC 836
               || | |  |  || |    |   || || || |||||| |||| | | |  || ||
Db     525 GCCCCCCCCCGCGCCCGCCGCCCCCCGCGCGGCGGGGCCCCGCGCCC-CGCCCGGCGCCC 583

Qy     837 ACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAACCA 896
           |||| ||  | | |  | || || | ||  || | |||   ||||   ||||    ||
Db     584 CCGGCGCCCGCGCCCGCCCGGCCCCCCCCCCCTCCCGGTCCGCGCCGCGGCTGCCCCG 643

Qy     897 CCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCGTGTCGGTCCGCGCTCGGCAA 956
           |||  ||  | | ||     | |  |||| || ||||   |||| || ||| ||
Db     644 CCCGGCGCCGCCCCGCGCCGGGCGGCGCCGGGCGGGCCCCGCTCGGGCCCCGCGCGCGGC 703

Qy     957 CGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGAC 995
           |  |  ||| | | | |||   || |      | |||| |
Db     704 CCGGCGGGCGCCCCCCGCTCGCTCGCTCNCGCCCCCGGC 742


RESULT 13
AQ251130
LOCUS       AQ251130                798 bp    DNA     linear   GSS 07-OCT-1998
DEFINITION  F26H6-Sp6 IGF Arabidopsis thaliana genomic clone F26H6, DNA
            sequence.
ACCESSION   AQ251130
```

```
VERSION     AQ251130.1  GI:3704196
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
            Ecker,J.
  TITLE     BAC End Sequences at ATGC
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Ecker J.
            Arabidopsis Thaliana Genome Center
            University of Pennsylvania
            Dept. of Biology, University of Pennsylvania, Philadelphia, PA
            19104
            Tel: 215-898-9384
            Fax: 215-898-8780
            Email: jecker@atgenome.bio.upenn.edu
            Seq primer: Sp6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .798
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="F26H6"
                     /clone_lib="IGF"
                     /sex="hermaphrodite"
                     /note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI;
                     Produced by Thomas Altmann"
BASE COUNT      347 a     99 c     90 g    259 t      3 others
ORIGIN

  Query Match             4.1%;  Score 54;  DB 12;  Length 798;
  Best Local Similarity   53.9%;  Pred. No. 0.37;
  Matches  132;  Conservative    0;  Mismatches  111;  Indels    2;  Gaps    1;

Qy     226 TGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACAT 285
           ||  |||| | || ||  |  |     |  |  | ||||  |  | | |   |     |
Db     152 TGATTATTGTAATAGTAATCAAAAAGGTGAATTAATAGTATTTAACATAATAATTAGTAA 211

Qy     286 ATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTT 345
           |  ||||  ||| | |   |  |||  ||| || | |||||| ||||  |     | |  |
Db     212 AAAATCATTTATTTGAATTTTGTTTATTAATTTCATTTAAAATTAAAGTTCTGCATAAAT 271

Qy     346 CTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTA 405
           |||  |  ||  | |||  |||  || ||| ||| |  ||||  |||| || | |     |
Db     272 NTAAAATAAATTATATTATAAATTTCTCTTCAAAACATTTATTTATTTTATAATAAAAAA 331

Qy     406 TTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGTACGACT 465
           ||   ||||  | |  | |   || |||   | || |  |||||| |  ||| ||| | |
Db     332 TT--AAATTATATTACAAATTTAAAATAAGTTTCACTTTAAAACATTTATTTATACTAAT 389

Qy     466 ATATA 470
           || |
Db     390 TTAAA 394


RESULT 14
CNS00FVE
LOCUS       CNS00FVE                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR32P18 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL071298
VERSION     AL071298.1  GI:4951138
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR32P18"
                     /note="end : TET3"
BASE COUNT      267 a    223 c     31 g    184 t    396 others
ORIGIN

  Query Match             4.1%;  Score 54;  DB 12;  Length 1101;
  Best Local Similarity   29.1%;  Pred. No. 0.37;
  Matches  102;  Conservative   89;  Mismatches  159;  Indels    0;  Gaps    0;

Qy     204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
           |: |:: : | |   |: :|: |  |::  :  : |      || |  | || | | :|
Db     551 TWAAWHAWWTMCAAATHTWTWATTAAWMMAYAAWATAAACCATAATTTTAATCCAAWWAM 610

Qy     264 TTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATT 323
           :| :: |::|:   :  |: ||  | | : ||:| : |  : |  |:  | :::|: ||:
Db     611 WTTHMWAWMCWCWAMACTWWATCCCTTTWAATHTTWTATTMAAAATWATTTWMWTWTATY 670

Qy     324 AAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACT 383
           |:|:   ::: |     |:::|  | |  :|  ::: | : |: |       |::  ||
Db     671 AHAWAATWMWTTCCWMTAMWWTTATAAAAWMAATWMWATAHTAWATAMATCTTAWYTTCT 730

Qy     384 TTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATG 443
           |  || || :| ||| |: :  ||  |   | | | : :||:  ::|:: :  :  ::
Db     731 ATCTCTTTTAYAWATTTTYTYYATTTCACAATAAWTTWCWAAMWTMMAWWTWTAMACMWW 790

Qy     444 GAAAACAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAG 503
           ||| |  |  :  |  |  ::  ::::  | :| | :   |  : :|: : ::| :::
Db     791 CAAATCTATTHAATYAATCWYATYWWHWWCTWTWTTHTHWTTTHTHAYCHTHMACWHWCW 850

Qy     504 TCGTTTGAAGATTGAACCTATATATCGGTACGGTTAATCCGTCTATGTAC 553
           || ||: || | |  |  |:::|  |  | :  :|:|:|: | :|| : |
Db     851 TCHTTYCAAWACTWTAAATMWWTTCCCCTCMTWWTHAYCMTTTWATCWTC 900


RESULT 15
CNS016IP/c
LOCUS       CNS016IP                481 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN16B22 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL106795
VERSION     AL106795.1  GI:5623841
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
```

```
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 481)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .481
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN16B22"
                     /note="end : T7"
BASE COUNT        0 a    113 c    142 g     67 t    159 others
ORIGIN

  Query Match             4.1%;  Score 53.8;  DB 12;  Length 481;
  Best Local Similarity   30.7%;  Pred. No. 0.39;
  Matches  138;  Conservative   50;  Mismatches  261;  Indels    0;  Gaps    0;

Qy     674 CAACAGGCAGGCAGGCAGGGACCAGGGGCCGCCTTTATTTCAGGCTCGCTGGCCCCACGG 733
           |  |  || |   ||||  :  | | : ||  |         ::|   |    |||::|:|
Db     473 CNNCGCGCCGMNCGGCACCSCNCNGCRCCCMACMMANNNNAMSGNGGGGNNNCCCSVCVG 414

Qy     734 GCGTGCTGCGTGCACGAAGGGCACTACCCCAACCTCTCACCGAAAAACCGCGCTGGATCG 793
           | || | :  : : :|    || :|::|: || | |  | ::|:
Db     413 CGGGGCNGNSNCMCMCMAMCMAACMMCMMCMCCCMCCCNACCMMAMNNNANNNSNNNNGM 354

Qy     794 GCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTCTG 853
           :       | | |   |||   | |  | | |  |  ||::  | :|   || |
Db     353 ANNMNMNGNCNNAAGCGCGNCCCGNNGNCNNCCCNCNCCCCCCVVCCCCSCCNNCCCCCC 294

Qy     854 CAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAACCACCCCCAACCCACTCACG 913
           | : : |      | |  |  | |  |:||   |||  |         |  |:: : : |:
Db     293 CNSNSCCCNNNNCGNCGSGGGGGGGGGSGGGCGGGCGGACGGCGGGGCCNCSSSSSSSCS 234

Qy     914 AAACCCCGTCCCGGCCGTGCCCGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTG 973
           |||:  :|::|: | | ||         | | |     |    :  ::|| | :  |
Db     233 SCCCCCSCCSCSSGSGGGGGCCCCNNNNCCNCACCCCNCMCCSNCNSAGSSCCNGNSGCG 174

Qy     974 CTGAGTCCCCTGGACACCCGACACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCTCA 1033
           :  || | |    |: ::: |    : |  |||     |   :   : |  | ||    :
Db     173 SGCAGCCGCGSAAASGSVVAANCGGVAGNNGGCNGSNCGGGSVRRARGCARGCAAGGGAR 114

Qy    1034 TCTGCCCCCACGGCCGACTGCGCTGCGCCGCCCGGATATATATACCCATCGTTATCGATC 1093
           | | | |    || |   | |  |    |  || |   | | |   | |   |   | |
Db     113 GCCGGCNCGNGAGCAGCARGGGGAGGANGGGACGCAACCAAAAAAAAACCCNGANGCAGC 54

Qy    1094 GATCGATCGCGTCACTCACGGGTAGCTCA 1122
           | ||:    | : |  | | |  || ||
Db      53 AAGCGMAGAGGCMGCANAAGNGGNGCNCA 25


Search completed: November  7, 2002, 12:25:20
Job time : 1257.97 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:47:38 ; Search time 1685.86 Seconds
(without alignments)
16248.621 Million cell updates/sec

Title: US-09-905-558C-16
Perfect score: 1309
Sequence: 1 cccatcgctgctttgtctac.....caagggaaagtgatcccatg 1309

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	232.8	17.8	2126	6	AX299951	AX299951 Sequence
c 2	124.6	9.5	1888	6	AX099709	AX099709 Sequence
c 3	112.2	8.6	9285	8	ZMU09989	U09989 Zea mays D3
4	60.2	4.6	174802	2	AC108209	AC108209 Homo sapi
5	58	4.4	125020	9	AF429315	AF429315 Homo sapi
6	56.6	4.3	173848	9	AC090043	AC090043 Homo sapi
7	56.2	4.3	169660	9	AL161449	AL161449 Human DNA
8	55.6	4.2	204652	2	PFMAL13P6	AL049183 Plasmodiu
c 9	55.4	4.2	65523	2	AC024266	AC024266 Homo sapi
10	55.2	4.2	650	3	SGU36751	U36751 Schizaphis
c 11	54.8	4.2	204652	2	PFMAL13P6	AL049183 Plasmodiu
12	54	4.1	89904	8	ATT25K17	AL049171 Arabidops
13	54	4.1	196286	8	ATCHRIV64	AL161564 Arabidops
c 14	53.8	4.1	115522	8	AC006248	AC006248 Arabidops
c 15	53.6	4.1	214033	2	AC105576	AC105576 Rattus no
16	53.4	4.1	614	3	AF254090	AF254090 Rhopalosi
c 17	53.2	4.1	167553	9	AC090042	AC090042 Homo sapi
18	53	4.0	127539	9	CNS01DSN	AL121821 Human chr
19	53	4.0	146868	2	AC016802	AC016802 Homo sapi
20	53	4.0	151117	2	AL357128	AL357128 Homo sapi
21	53	4.0	161277	2	AC091953	AC091953 Homo sapi
22	53	4.0	176033	2	AL161794	AL161794 Homo sapi
23	53	4.0	178342	9	AC007435	AC007435 Homo sapi
24	53	4.0	187150	2	AC025924	AC025924 Homo sapi
c 25	53	4.0	199385	2	AC019049	AC019049 Homo sapi
c 26	52.6	4.0	72090	2	AC095644	AC095644 Rattus no
27	52.6	4.0	80692	2	AC100569	AC100569 Mus muscu
28	52.6	4.0	110000	2	PFMAL4P1_0	AL034557 Plasmodiu
29	52.6	4.0	141532	2	AC094940	AC094940 Rattus no
30	52.6	4.0	191772	2	AC078975	AC078975 Homo sapi
31	52.6	4.0	293431	2	PFMAL13P4	AL049181 Plasmodiu
32	52.4	4.0	171574	2	AC012300	AC012300 Homo sapi
c 33	52.4	4.0	196203	9	AL355355	AL355355 Human DNA
c 34	52.2	4.0	90487	9	AL592166	AL592166 Human DNA
c 35	52.2	4.0	126801	9	AC004849	AC004849 Homo sapi
c 36	52.2	4.0	149450	2	AC067880	AC067880 Homo sapi
37	52.2	4.0	153300	9	AC073387	AC073387 Homo sapi
c 38	52.2	4.0	170141	2	AL611928	AL611928 Homo sapi
39	52.2	4.0	213692	2	AL590427	AL590427 Homo sapi
c 40	52	4.0	159593	9	AC004832	AC004832 Homo sapi
41	51.8	4.0	678	3	AF143595	AF143595 Myzus var
c 42	51.8	4.0	14135	3	AE001388	AE001388 Plasmodiu
c 43	51.6	3.9	79646	8	ATT10D17	AL353865 Arabidops
44	51.6	3.9	162996	9	AC006441	AC006441 Homo sapi
c 45	51.4	3.9	64148	2	AC101402	AC101402 Mus muscu

ALIGNMENTS

RESULT 1
AX299951
LOCUS AX299951 2126 bp DNA linear PAT 26-NOV-2001
DEFINITION Sequence 92 from Patent WO0183790.
ACCESSION AX299951
VERSION AX299951.1 GI:17129442
KEYWORDS .
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (sites)
AUTHORS Conner,T.W., Dubois,P., Malven,M. and Masucci,J.D.
TITLE Plant regulatory sequences for selective control of gene expression
JOURNAL Patent: WO 0183790-A 92 08-NOV-2001;
Monsanto Technology LLC (US)
FEATURES Location/Qualifiers
source 1. .2126
/organism="Zea mays"
/db_xref="taxon:4577"
BASE COUNT 607 a 419 c 433 g 667 t
ORIGIN

```
  Query Match             17.8%;  Score 232.8;  DB 6;  Length 2126;
  Best Local Similarity   83.0%;  Pred. No. 2.7e-44;
  Matches  289;  Conservative    0;  Mismatches   57;  Indels    2;  Gaps    2;

Qy       5 TCGCTGCTTTGTCTACATCATGTTCTTCATCATCCTCCCCAGGCGACGCGTGCTGCTGTT 64
           ||||    | ||| ||| | || | ||| ||  | |||| ||  ||||||||||||||||
Db      82 TCGCGTGATCGTCCACAGCTTGCTGTTCGTCGCCTTCCCAAGTTGACGCGTGCTGCTGTT 141

Qy      65 CTTATTCA-GACTACCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGT 123
           ||| |||  | | |||||||||| |||||||  |||||||||| ||||| |||||| ||
Db     142 CTTCTTCCCGGCGACCGTTCGAGGGACTGCACTGCGTACATCTTCCTGCACCGACTTCGT 201

Qy     124 ACGGCTACATCGAACATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAA 183
           |||||||||||||||| | ||||||||||||||||||||| ||| ||||  |||||| |
Db     202 ACGGCTACATCGAACAAACACACGAGATGTCTCGTGTGAATGGAGCCACTGGTGCCTTGA 261

Qy     184 GCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTG 243
           ||||||||  |||||  ||||||| |||||||||| ||||||||||  ||| |||| ||
Db     262 GCATCGGTCCCTCCGCTGGGTACACTCTGTTCTTCGTATTTGTGCAT-GTTTCATTGCTG 320

Qy     244 TTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACA 303
           ||||||| |||| ||||||||||||| | |||||||||||| ||||||||||||| ||
Db     321 TTTACTGCTTATGCGAGTAGTTATACACATATGCACATACATGTCATCACATATATCGCA 380

Qy     304 ATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACA 351
           |  || |||  |||||||||||||||||||| |||| |||||||||
Db     381 CTGATTATCTGGATTAAATTAAAACTAAAAATGCCTAACTTTCTAACA 428


RESULT 2
AX099709/c
LOCUS       AX099709                1888 bp    DNA     linear   PAT 02-APR-2001
DEFINITION  Sequence 65 from Patent WO0119976.
ACCESSION   AX099709
VERSION     AX099709.1  GI:13538763
KEYWORDS    .
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 1888)
  AUTHORS   Anderson,H.M., Chay,C.A., Chen,G. and Conner,T.W.
  TITLE     Plant regulatory sequences for control of gene expression
  JOURNAL   Patent: WO 0119976-A 65 22-MAR-2001;
            MONSANTO COMPANY (US)
FEATURES             Location/Qualifiers
     source          1. .1888
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
BASE COUNT      569 a    443 c    423 g    453 t
ORIGIN

  Query Match              9.5%;  Score 124.6;  DB 6;  Length 1888;
  Best Local Similarity   63.9%;  Pred. No. 8.4e-19;
  Matches  227;  Conservative    0;  Mismatches  114;  Indels   14;  Gaps    2;

Qy      78 CCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGTACGGCTACATCGAA 137
           | ||| ||| |||||||  |||||  | ||  |||||||||  || ||||   || |||||||
Db     784 CTGTTTGAGGGACTGCACTGCGTAGACCTGCCTGCATCGAGTTCGTACAATTATATCGAA 725

Qy     138 CATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCC 197
           ||||  |||||||||||| ||||||||  | | | || || | || |||||||||  || |
Db     724 CATACACACGAGATGTCTTGTGTGAATTGGGCCGCTGATACTTTGAGCATCGGTCCCTTC 665

Qy     198 GTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATAC 257
             |||| || | ||||| | | || || |       | |  || |||| ||||| ||||||
Db     664 ACAGGGAACTTCCTGTTTTCCATACTTTT--TGCGTGTCTTTTTTGTGTACTGCTTATAC 607

Qy     258 GAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAAT 317
           || ||||  |            |||||| || ||||||| || |||||  ||||||  |
Db     606 GAATAGTGGT-----------ATACATGTCGTCACATAAATTACAATGGTTTTCTGGAC 559

Qy     318 TAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCA 377
           ||||||||||||  |||| || | |||||||||       ||| |   |    ||    |
Db     558 TAAATTAAAACTAGTAATGCCTCATTTTCTAACATATTTCACAGTAATAAAAGTTAGGTA 499

Qy     378 ACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACAT 432
             || ||    || |  |      | | ||  |      | | || |||  |  | ||
Db     498 TCAGCTAATTCTTTGGTGATAGGGTATAAATGACGATGCTCTATATTTATCAGAT 444


RESULT 3
ZMU09989/c
LOCUS       ZMU09989                9285 bp    DNA     linear   PLN 06-DEC-2000
DEFINITION  Zea mays D3L H(+)-transporting ATPase (Mha1) gene, complete cds.
ACCESSION   U09989
VERSION     U09989.1  GI:507770
KEYWORDS    .
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 2391 to 2453; 2525 to 2659; 2751 to 2930; 3019 to 3225;
            3306 to 3425; 3510 to 3629; 3722 to 3844; 3927 to 3938)
  AUTHORS   Jin,Y. and Bennetzen,J.L.
  TITLE     Integration and Nonrandom Mutation of a Plasma Membrane Proton
            ATPase Gene Fragment within the Bs1 Retroelement of Maize
  JOURNAL   Plant Cell 6, 1177-1186 (1994)
  MEDLINE   95003707
REFERENCE   2  (bases 1 to 9285)
  AUTHORS   Jin,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAY-1994) Young-Kwan Jin, Biological Sciences, Purdue
            University, 339 Hansen Life Science Research Building, West
            Lafayette, IN 47907, USA
FEATURES             Location/Qualifiers
     source          1. .9285
                     /organism="Zea mays"
                     /strain="D3L"
                     /db_xref="taxon:4577"
                     /chromosome="2L"
                     /map="2L"
                     /clone="C18"
                     /clone_lib="Sau3AI partial of maize genomic DNA into
                     Charon  40"
     mRNA            join(<1446. .1508,1874. .1993,2105. .2203,2316. .2453,
                     2525. .2659,2751. .2930,3019. .3225,3306. .3425,3510. .3629,
                     3722. .3844,3927. .4031,4113. .4257,4379. .4543,4637. .4875,
                     4964. .4996,5091. .5172,5257. .5417,5513. .5689,5768. .5950,
                     6036. .>6290)
                     /gene="Mha1"
                     /product="H(+)-transporting ATPase"
     gene            join(<1446. .1508,1874. .1993,2105. .2203,2316. .2453,
                     2525. .2659,2751. .2930,3019. .3225,3306. .3425,3510. .3629,
                     3722. .3844,3927. .4031,4113. .4257,4379. .4543,4637. .4875,
                     4964. .4996,5091. .5172,5257. .5417,5513. .5689,5768. .5950,
                     6036. .>6290)
                     /gene="Mha1"
     exon            <1446. .1508
                     /gene="Mha1"
                     /number=1
     CDS             join(1446. .1508,1874. .1993,2105. .2203,2316. .2453,
                     2525. .2659,2751. .2930,3019. .3225,3306. .3425,3510. .3629,
                     3722. .3844,3927. .4031,4113. .4257,4379. .4543,4637. .4875,
                     4964. .4996,5091. .5172,5257. .5417,5513. .5689,5768. .5950,
                     6036. .6290)
                     /gene="Mha1"
                     /codon_start=1
                     /product="H(+)-transporting ATPase"
                     /protein_id="AAB60276.1"
                     /db_xref="GI:533775"
                     /translation="MGTTEDKASNLDAVLKEAVDLENIAIQEVFESLRCSPQGLSTEQ
                     AEQRLAIFGPNKLEEKQESKFLKFLGFMWNPLSWVMEAAAIMAIALANGGNKPPDWQD
```

```
                FVGIITLLIINSTISFIEENNAGNAAAALMARLAPKAKVLRNGRWAEEESAILVPGDI
                ISVKLGDIIPADARLLEGDPLKIDQSALTGESLPVTKGPGDGVYSGSTCKQGEIEAVV
                IATGVHTFFGKAAHLVDSTNQVGHFQKVLTAIGNFCICSIAVGMLVEIIVMYPIQHRA
                YRSGIDNLLVLLIGGIPIAMPTVLSVTMAIGAHRLAQQGAITKRMTAIEEMAGMDILC
                SDKTGTLTLNQLTVDKSLVEVFQRGVDQDTVILMAARASRTENQDAIDATIVGMLADP
                TEARAGVQEIHFLPFNPTDKRTALTYLDGEGRMHRVSKGAPEQILHLAHNKKDIETRV
                RAVIDNFAERGLRALGVAYQEVPDGRKESPGGPWEFMGLLPLFDPPRKDSADTISKAL
                DLGVNVKMITGDQLAIGKETARRLGMGTNMYPSSALLEQNKDESIASLPIDELIETAD
                GFAGVFPEHKYEIVKRLQARKHISGMTGDGVNDAPALKKADIGIAVADSTDAARSASD
                IVLTEVGLSVIISAVLTSRAIFQRMKNYTIYAVSITIRIVLGFMLLALIWKFDFPPFM
                VLIIAILNDGTIMTISKDRVKPSPQPDSWKLAEIFATGVVLGTYLAMMTVIFFWAAYK
                TDFFPRLFHVESLAHDDFQMLASAVYLQVSTISQALIFVTRSRSWSFVERPGFLLVSA
                FLVAQLIATLIAVYANWAFAAIKGIGWGWAGVIWLYNIVCYLPLDIIKFLIRYALSGR
                AWNLVLEQRIAFTSKKNFGAEERERRWAHAQRSLHGLELQPPEAASMFENKTSFSEVN
                QLAEEARRRAEMARSDCCLSSISTERGDQTLHCPV"
intron          1509. .1873
                /gene="Mha1"
                /number=1
exon            1874. .1993
                /gene="Mha1"
                /number=2
intron          1994. .2104
                /gene="Mha1"
                /number=2
exon            2105. .2203
                /gene="Mha1"
                /number=3
intron          2204. .2315
                /gene="Mha1"
                /number=3
exon            2316. .2453
                /gene="Mha1"
                /number=4
intron          2454. .2524
                /gene="Mha1"
                /number=4
exon            2525. .2659
                /gene="Mha1"
                /number=5
intron          2560. .2750
                /gene="Mha1"
                /number=5
exon            2751. .2930
                /gene="Mha1"
                /number=6
intron          2931. .3018
                /gene="Mha1"
                /cons_splice=(5'site:no, 3'site:yes)
                /number=6
exon            3019. .3225
                /gene="Mha1"
                /number=7
intron          3226. .3305
                /gene="Mha1"
                /number=7
exon            3306. .3425
                /gene="Mha1"
                /number=8
intron          3426. .3509
                /gene="Mha1"
                /number=8
exon            3510. .3629
                /gene="Mha1"
                /number=9
intron          3630. .3721
                /gene="Mha1"
                /number=9
exon            3722. .3844
                /gene="Mha1"
                /number=10
intron          3845. .3926
                /gene="Mha1"
                /number=10
exon            3927. .4031
                /gene="Mha1"
                /number=11
intron          4032. .4112
                /gene="Mha1"
                /number=11
exon            4113. .4257
                /gene="Mha1"
                /number=12
intron          4258. .4378
                /gene="Mha1"
                /number=12
exon            4379. .4543
                /gene="Mha1"
                /number=13
intron          4544. .4636
                /gene="Mha1"
                /number=13
exon            4637. .4875
                /gene="Mha1"
                /number=14
intron          4876. .4963
                /gene="Mha1"
                /number=14
exon            4964. .4996
                /gene="Mha1"
                /number=15
intron          4997. .5090
                /gene="Mha1"
                /number=15
exon            5091. .5172
                /gene="Mha1"
                /number=16
intron          5173. .5256
                /gene="Mha1"
                /number=16
exon            5257. .5417
                /gene="Mha1"
                /number=17
intron          5418. .5512
                /gene="Mha1"
                /number=17
exon            5513. .5689
                /gene="Mha1"
                /number=18
intron          5690. .5767
                /gene="Mha1"
                /number=18
exon            5768. .5950
                /gene="Mha1"
                /number=19
intron          5951. .6035
                /gene="Mha1"
                /number=19
exon            6036. .one-of(6408,6409)
                /number=20
                /evidence=experimental
polyA_site      6408
                /evidence=experimental
polyA_site      6409
                /evidence=experimental
BASE COUNT     2329 a   2181 c   2086 g   2689 t
ORIGIN

  Query Match             8.6%;  Score 112.2;  DB 8;  Length 9285;
  Best Local Similarity   74.1%;  Pred. No. 9.8e-16;
  Matches  177;  Conservative    0;  Mismatches   43;  Indels   19;  Gaps    2;

Qy      78 CCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGTACGGCTACATCGAA 137
           ||||||||| ||||||| | ||||| || |||||||||| |||||| | ||||||||||
Db     220 CCGTTCGAGGGACTGCACTGTGTACACCTGTCTGCATCGAATTTGTATGACTACATCGAA 161

Qy     138 CATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCC 197
```

```
          |||| |||||||||||||| ||||||| ||| |||| |||                    |
Db    160 CATACACACGAGATGTCTCATGTGAATGGAGCCACTGATG------------------TC 119

Qy    198 GTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATAC 257
          | ||||||| ||    | | | || |||||||  ||| |||  ||||||||  |||| |
Db    118 GCAGGGTACTTTTAAATTTCCATACTTGTGCAT-GTTTCATTTATGTTTACTACTTATGC 60

Qy    258 GAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAA 316
          ||||||  |||| |||| |||||||| ||||||||||||||||  | || ||||| |
Db     59 CAGTAGTGGTACACACATACACATACATGTCATCACATATATCACGTTGTTCTTCTAGA 1


RESULT 4
AC108209
LOCUS       AC108209              174802 bp    DNA     linear   HTG 31-JAN-2002
DEFINITION  Homo sapiens chromosome 4 clone RP11-439B3, WORKING DRAFT SEQUENCE,
            7 unordered pieces.
ACCESSION   AC108209
VERSION     AC108209.2  GI:18450201
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174802)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174802)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jan 31, 2002 this sequence version replaced gi:18376960.

            ------------------- Genome Center --------------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            Contact: submissions@watson.wustl.edu
            ---------------- Project Information -----------------
            Center project name: H_NH0439B03
            ------------------------------------------------------
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 2%
            Sequencing vector: plasmid; 98%
            Chemistry: Dye-primer ET; 0% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 173471 bases at least Q40
            Consensus quality: 173768 bases at least Q30
            Consensus quality: 173922 bases at least Q20
            Insert size: 173000; agarose-fp
            Insert size: 174202; sum-of-contigs
            Quality coverage: 12.07 in Q20 bases; agarose-fp
            Quality coverage: 12.00 in Q20 bases; sum-of-contigs
            ------------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1162: contig of 1162 bp in length
            *     1163     1262: gap of unknown length
            *     1263     2346: contig of 1084 bp in length
            *     2347     2446: gap of unknown length
            *     2447     8520: contig of 6074 bp in length
            *     8521     8620: gap of unknown length
            *     8621    24285: contig of 15665 bp in length
            *    24286    24385: gap of unknown length
            *    24386    63017: contig of 38632 bp in length
            *    63018    63117: gap of unknown length
            *    63118   113781: contig of 50664 bp in length
            *   113782   113881: gap of unknown length
            *   113882   174802: contig of 60921 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174802
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /clone="RP11-439B3"
     misc_feature    1. .1162
                     /note="assembly_name:Contig23"
     misc_feature    1263. .2346
                     /note="assembly_name:Contig27"
     misc_feature    2447. .8520
                     /note="assembly_name:Contig32"
     misc_feature    8621. .24285
                     /note="assembly_name:Contig33"
     misc_feature    24386. .63017
                     /note="assembly_name:Contig34
                     clone_end:T7
                     vector_side:left"
     misc_feature    63118. .113781
                     /note="assembly_name:Contig35
                     clone_end:SP6
                     vector_side:right"
     misc_feature    113882. .174802
                     /note="assembly_name:Contig36"
BASE COUNT    54963 a  33646 c  34740 g  50806 t    647 others
ORIGIN

  Query Match             4.6%;  Score 60.2;  DB 2;  Length 174802;
  Best Local Similarity   52.6%;  Pred. No. 0.0033;
  Matches  131;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy    204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
          || | | | ||  | | |||     |   | |  || |   | | | | |||     ||
Db  58325 TATAATATATTTATATAATTATATAAATGTATACTTATATATAATTCAATATTTATATAA 58384

Qy    264 TTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATT 323
          |||||| |||  |  | ||  |||| || | |||||  ||||||||| | ||| |  |  |
Db  58385 TTATATATAATTCAATATTTATATAATTATATATAATTCAATATTTATATAATTATATAT 58444

Qy    324 AAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACT 383
          ||  | | |  |   ||| | | ||       | | |||   ||  | | |  |||| |
Db  58445 AATTCAATATTTATATAATTATATATATAATTCAATATTTATATAATATATTTAACATAT 58504

Qy    384 TTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATG 443
          |   |||  |    | || ||| |    |||   | ||||||  |   ||||| ||
Db  58505 ATTTATATAATTATATATTATATATAATTTTTGTATATATAAAATATTCAGTCTTAAAAA 58564

Qy    444 GAAAACAGT 452
          ||| ||| |
Db  58565 GAAGACAAT 58573


RESULT 5
AF429315
LOCUS       AF429315              125020 bp    DNA     linear   PRI 18-JAN-2002
DEFINITION  Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION   AF429315
VERSION     AF429315.1  GI:17646244
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
            Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
            Potter,N.T., Ross,C.A. and Margolis,R.L.
```

```
  TITLE     A repeat expansion in the gene encoding junctophilin-3 is
            associated with Huntington disease-like 2
  JOURNAL   Nat. Genet. 29 (4), 377-378 (2001)
  MEDLINE   21583737
   PUBMED   11694876
REFERENCE   2  (bases 1 to 125020)
  AUTHORS   Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
            Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
FEATURES             Location/Qualifiers
     source          1. .125020
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /map="16q24.3; between D16S520 and WI-12410"
                     /note="isolated from a patient with Huntington's
                     Disease-Like 2 (HDL2)"
     repeat_region   complement(35581. .35746)
                     /rpt_type=tandem
                     /rpt_unit=ctg
     mRNA            complement(<36507. .>36887)
                     /gene="JPH3"
                     /product="junctophilin 3"
     gene            complement(<36507. .>36887)
                     /gene="JPH3"
                     /note="JP3"
     CDS             complement(<36507. .36887)
                     /gene="JPH3"
                     /note="component of the junctional complex between plasma
                     membrane and endoplasmic reticulum"
                     /codon_start=1
                     /product="junctophilin 3"
                     /protein_id="AAL40941.1"
                     /db_xref="GI:17646245"
                     /translation="MSSGGRFNFDDGGSYCGGWEDGKAHGHGVCTGPKGQGEYTGSWS
                     HGFEVLGVYTWPSGNTYQGTWAQGKRHGIGLESKGKWVYKGEWTHGFKGRYGVRECAG
                     NGAKYEGTWSNGLQDGYGTETYSDG"
BASE COUNT    29056 a  32731 c  30696 g  28283 t   4254 others
ORIGIN

  Query Match             4.4%;  Score 58;  DB 9;  Length 125020;
  Best Local Similarity   11.8%;  Pred. No. 0.01;
  Matches   70;  Conservative  248;  Mismatches  273;  Indels    0;  Gaps    0;

Qy     699 GGGCCGCCTTTATTTCAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACT 758
           ||| | |   || |   |||   |      :: :   : ::  :::: ::  ::::: ::
Db   16847 GGGACCCAGGTAGTAGCGGCCATCGCARRKMMKSHAGRRRMCYYSKSWMSMBMSVSYSVK 16906

Qy     759 ACCCCAACCTCTCACCGAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGT 818
           ::::::| :: ::   |  :  |::: :::::: ::  :    :: : | :   : :: ::
Db   16907 MHSMHASBSCMHWBKCMTWSCCMSMMYKSSWWGSSWGMCCWGRRRSKGWKWYSRGMSRSK 16966

Qy     819 GCCCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTC 878
           :: :   :::::::: :| :: :|::  :||::  |:::| | |:::  |::: |  : ::
Db   16967 SMRYTGGSKMMRSSMMCTSSCYASMCCMCCWSCCCMRSCCCMCRSYCCCMRYCCACKCYM 17026

Qy     879 GCGGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCGTGT 938
           ::   :: ::::::::::  ::  :: : : ::::   :  || :: :|  |:::  :|:
Db   17027 SSYWTMSASYMSRRSYWKRSMKCWSRMCSRSRSSKCKSRGCGGSMGKGKGGKSYGRKTKR 17086

Qy     939 CGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCC 998
           ::  ::   ::  :::: ::    :: |: |  |: :|   |: : :   :   | :|:
Db   17087 KSRGKMKGAKWYMYYRRSMRKMKMYSSKGMYCMYCWCWGRRGCYCSCMTSRSAMCCSYCY 17146

Qy     999 TGTCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCACGGCCGACTGCGCTGC 1058
           :  : |:: ::|     ::  |:::  : :::   : :: :    |    | | |
Db   17147 AKCKSMCYSCYYGMSMKGYYYACSYRGSMSSKYCMRGSTYSTSCGCCCTTTTTCCCCCNA 17206

Qy    1059 GCCGCCCGGATATATATACCCATCGTTATCGATCGATCGATCGCGTCACTCACGGGTAGC 1118
             |   | | |       : ::::   | :     :   :  : |:  : | : : ::
Db   17207 NTGGGGAAGCTTTTNCNKKTYSYYRKRNGCAMCKYNNYYNNSWRSSCRAGMSCTKYKSSM 17266

Qy    1119 TCATGGTCGAGCGTAGCATGCAGGAACTTATTTGCCGTGCGCTCCCAGGTCTCCGCTCGC 1178
           |:   ::|   : :  :: ::|:: :::  : : : :    : ::     :::| | :::
Db   17267 TMSMASYCWCMSMYYCSMRSMASRGMSWSSYMKMMKSSWRMSCYMKCCWCMKMCYCMMRS 17326

Qy    1179 GTGCCTTCCAGTCTGTCTCACACTAGCTGCTGTGGGACGATCGAAGTGGGTGTGTCAGCT 1238
             :   :::  ::: :   | ||: :: : :::   :   : : :  :| ::: |:|:
Db   17327 MRSRGMSYYMYASWKSSSRGCYCTRCYWCMSSKSCYKSYYMMMRSKRMKGMKWSRCWGSM 17386

Qy    1239 AGCTAGCTGCGCCGTGACCACGCACATGACCGCAGTGCGCGCGGGGCTGAT 1289
           :|:  ::  ::|::  :|::::: |   | | ::: |:   ::  ::::::
Db   17387 WGSASRSSCYKCYKSMRCSMMSSKCYRCAGCMMMKGGYMRYMRCWSMKRRW 17437


RESULT 6
AC090043
LOCUS       AC090043              173848 bp    DNA     linear   PRI 11-FEB-2001
DEFINITION  Homo sapiens chromosome 3 clone RP11-551L4 map 3p, complete
            sequence.
ACCESSION   AC090043
VERSION     AC090043.1  GI:12745081
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173848)
  AUTHORS   Xiong,H., Zhou,Y., Dong,H., Lin,W., Chen,B., Zhang,C., Zhang,Y.,
            Cai,Z., Ying,H.F., Wang,H., Gu,W., Zhu,G., Tu,Y., Zhang,X., Jia,J.,
            Shen,H., Zhang,D., Wu,C., Lu,G., Zhong,M., Jiang,H., Ren,S., Fu,G.,
            Chen,Z. and Huang,M.
  TITLE     Chromosome 3p genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173848)
  AUTHORS   Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,
            Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D., Guo,Z.,
            He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,F.,
            Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y.,
            Li,W., Li,W., Li,Y., Luo,C., Luo,J., Niu,Y., Qi,Q., Qi,X., Song,L.,
            Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
            Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
            Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Yan,C., Yang,X.,
            Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
            Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
            Yu,J. and Yang,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2001) Human Genomic Center, Institute of
            Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
            100101, P.R.China
COMMENT     --------------Genome Center
            Center:Beijing Center
            Center code:Beijing
            Website:http://hgc.igtp.ac.cn
            http://www.genomics.org.cn
            Contact:hgc@igtp.ac.cn
            -------------- Project Information
            Center project name:1% project
            Center clone name: RP11-551L4
            -------------- Summary Statistics
            Sequencing vector: pUC18; 100% of reads
            Chemistry: Dye-terminator: ET 55% of reads
            Chemistry: Dye-terminator Big Dye; 45% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 501 bases at least Q40
            Consensus quality: 829 bases at least Q30
            Consensus quality: 1046 bases at least Q20
            Insert size: 1157; sum-of-contigs
            Quality coverage:   1.35x in Q20 bases;sum-of-contigs
            --------------
FEATURES             Location/Qualifiers
     source          1. .173848
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
```

```
                     /map="3p"
                     /clone="RP11-551L4"
BASE COUNT    56663 a  31451 c  31882 g  53852 t
ORIGIN

  Query Match             4.3%;  Score 56.6;  DB 9;  Length 173848;
  Best Local Similarity   47.0%;  Pred. No. 0.023;
  Matches  208;  Conservative    0;  Mismatches  234;  Indels    1;  Gaps    1;

Qy     100 TACATCTTTCTGCATCGACTTTGTACGGCTACATCGAACATATACACGAGATGTCTCGTG 159
           || |  ||| |  ||  | | | || |  |||||  |   |       | || | | ||
Db  163709 TATAAATTTATATATATTATATATGTATACATATATTTATATATAAATTTATAGTT 163768

Qy     160 TGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCT 219
           |  ||| |   |   ||   |          |||     |    ||  |    ||
Db  163769 TACATATAAATAGATATATTTATTTATATATTTATATATTTATATATTTATATTTTAATA 163828

Qy     220 TATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 279
           | | | |  |||||| ||| ||| | ||   |||||| |  ||  |||| |||||| |
Db  163829 TGTATTTATATATTTATATAGTTATATATAAATTATATGTATAAATATATATACATATAT 163888

Qy     280 ATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACT 339
           ||  ||||   ||||||||    |||||| |  | ||  |||  |   || | |   |
Db  163889 ATGGATATTTATACATATATATTTATATTTATACATATATAATATATTATATATTTATAT 163948

Qy     340 AAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTAT-TCTACAT 398
             ||  ||      |    || || || |   | |     | ||  |||   |||||
Db  163949 TTATACATATGTATGATATATTTATATTGATACATATATATTATATATTTATAAATACAT 164008

Qy     399 TGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTG 458
           | ||   ||     | | | | | || | |      | ||||     ||| |    | |
Db  164009 TATTTATTTATATGTATAAATATTTATATATGTATATATACATATATAAATATATGTATA 164068

Qy     459 TACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAGTCGTTTGAAGATTGA 518
           |     ||||      || |||   || |     | |  ||||| | | |  | |||
Db  164069 TGTACATATAATTTATATATGTACATATACATATATAATATATATACATATATAAATTAT 164128

Qy     519 ACCTATATATCGGTACGGTTAAT 541
           |  || ||||    |    ||||
Db  164129 ATGTACATATACACATATATAAT 164151


RESULT 7
AL161449
LOCUS       AL161449              169660 bp    DNA     linear   PRI 04-DEC-2001
DEFINITION  Human DNA sequence from clone RP11-382H24 on chromosome 9p22.1-23
            Contains a thioredoxin peroxidase pseudogene, a SSB (Sjogren
            syndrome antigen B (autoantigen La) pseudogene, the 3' end of the
            MPDZ gene for multiple PDZ domain protein and a CpG island,
            complete sequence.
ACCESSION   AL161449
VERSION     AL161449.7  GI:8894259
KEYWORDS    HTG; CpG island; MPDZ.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169660)
  AUTHORS   Sehra,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 1, 2000 this sequence version replaced gi:8653807.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            This sequence is the entire insert of clone RP11-382H24 The true
            left end of clone RP11-664D14 is at 135950 in this sequence. The
            true right end of clone RP11-187K14 is at 53851 in this sequence.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP11-382H24 is from
            the library RPCI-11.2 constructed by the group of Pieter de Jong.
            For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1. .169660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="p22.1-23"
                     /clone="RP11-382H24"
                     /clone_lib="RPCI-11.2"
     misc_feature    2. .492
                     /note="match: GSS: Em:AQ700717"
     misc_feature    complement(1479. .2009)
                     /note="match: GSS: Em:AQ569033"
     repeat_region   1589. .1778
                     /note="MIR repeat: matches 23. .192 of consensus"
     repeat_region   1669. .1724
                     /note="2 copies 28 mer 92% conserved"
     repeat_region   2068. .2401
                     /note="L1MC4 repeat: matches 7453. .7834 of consensus"
     repeat_region   2409. .2828
                     /note="15 copies 28 mer 62% conserved"
     repeat_region   2429. .2848
                     /note="7 copies 60 mer 65% conserved"
     repeat_region   2452. .2801
                     /note="175 copies 2 mer ta 61% conserved"
     repeat_region   2819. .2911
                     /note="L1MC4 repeat: matches 7864. .7961 of consensus"
     misc_feature    complement(4102. .4478)
                     /note="match: GSS: Em:AQ180899"
     misc_feature    4497. .5002
                     /note="match: GSS: Em:AQ716496"
     repeat_region   4621. .4865
                     /note="MIR repeat: matches 1. .261 of consensus"
     repeat_region   5200. .5294
                     /note="L2 repeat: matches 2578. .2670 of consensus"
     repeat_region   6031. .6654
                     /note="L1PA12 repeat: matches 5583. .6152 of consensus"
     misc_feature    7419. .7966
                     /note="match: GSS: Em:AQ378047"
     repeat_region   7431. .7737
                     /note="AluSc repeat: matches 1. .307 of consensus"
     repeat_region   7763. .8103
                     /note="MLT1A1 repeat: matches 205. .536 of consensus"
     repeat_region   8379. .8422
                     /note="22 copies 2 mer tg 86% conserved"
     repeat_region   9171. .9447
                     /note="L1M4 repeat: matches 2900. .3176 of consensus"
     repeat_region   9450. .9739
                     /note="L1M4 repeat: matches 3869. .4175 of consensus"
     repeat_region   9774. .9835
                     /note="L1MA7 repeat: matches 6223. .6284 of consensus
                     L1MA7 repeat: matches 6223. .6284 of consensus"
     repeat_region   10099. .10533
                     /note="MLT1C repeat: matches 9. .466 of consensus"
```

```
repeat_region   10789. .11105
                /note="AluSx repeat: matches 1. .309 of consensus"
repeat_region   11109. .11267
                /note="MER5B repeat: matches 1. .178 of consensus"
misc_feature    complement(12290. .12569)
                /note="match: STS: Em:G05562"
gene            complement(12572. .13167)
                /gene="bA382H24.1"
                /pseudo
CDS             complement(12572. .13167)
                /gene="bA382H24.1"
                /note="bA382H24.1 (thioredoxin peroxidase pseudogene)
                match: cDNAs: Em:U27125 Em:AF221841
                match: proteins: Sw:Q91191 Tr:O93241 Sw:Q06830 Sw:P32119
                Sw:Q63716 Sw:P35704 Sw:Q90384 Tr:O44366 Sw:P35700
                Sw:Q61171 Tr:O96763 Tr:O61000 Tr:O85761"
                /codon_start=1
                /pseudo
                /evidence=not_experimental
misc_feature    14831. .15165
                /note="match: STS: Em:HS260YA5"
repeat_region   14993. .15030
                /note="19 copies 2 mer ac 97% conserved"
repeat_region   16682. .16772
                /note="MLT2D repeat: matches 1. .90 of consensus"
repeat_region   16759. .16999
                /note="MLT1A1 repeat: matches 1. .241 of consensus"
repeat_region   17339. .17638
                /note="AluY repeat: matches 4. .304 of consensus"
repeat_region   17645. .18235
                /note="L2 repeat: matches 1961. .2750 of consensus"
repeat_region   18526. .18676
                /note="MLT1J repeat: matches 33. .189 of consensus"
repeat_region   18882. .18943
                /note="MLT1J repeat: matches 411. .467 of consensus"
repeat_region   18986. .19291
                /note="AluSx repeat: matches 1. .307 of consensus"
repeat_region   19450. .19635
                /note="MIR repeat: matches 59. .248 of consensus"
repeat_region   20230. .20295
                /note="33 copies 2 mer at 75% conserved"
misc_feature    complement(21475. .21909)
                /note="match: GSS: Em:B46263"
repeat_region   21640. .22131
                /note="MLT1H repeat: matches 40. .547 of consensus"
repeat_region   22141. .22256
                /note="MLT1H repeat: matches 236. .352 of consensus"
repeat_region   22708. .22940
                /note="L1PA2 repeat: matches 5403. .5632 of consensus
                L1PA3 repeat: matches 5401. .5632 of consensus"
repeat_region   22941. .23434
                /note="L1PA2 repeat: matches 5641. .6144 of consensus"
repeat_region   22941. .23436
                /note="L1PA3 repeat: matches 5641. .6146 of consensus"
misc_feature    24368. .24949
                /note="match: GSS: Em:AQ744132"
misc_feature    24381. .24772
                /note="match: GSS: Em:AQ333224"
repeat_region   26193. .26290
                /note="MIR repeat: matches 97. .193 of consensus"
misc_feature    complement(26637. .27057)
                /note="match: GSS: Em:B75037"
repeat_region   27837. .27899
                /note="L2 repeat: matches 2641. .2708 of consensus"
repeat_region   28653. .28739
                /note="AluJo/FRAM repeat: matches 202. .289 of consensus"
repeat_region   29135. .29237
                /note="MIR repeat: matches 97. .194 of consensus"
repeat_region   29262. .29957
                /note="L2 repeat: matches 1986. .2742 of consensus"
repeat_region   29958. .30391
                /note="MLT2B repeat: matches 1. .444 of consensus
                MLT2B repeat: matches 1. .444 of consensus"
repeat_region   30392. .30590
                /note="L2 repeat: matches 1771. .1986 of consensus
                HERVL repeat: matches 5456. .5654 of consensus"
repeat_region   30591. .30631
                /note="MLT2B repeat: matches 1. .41 of consensus
                MLT2A repeat: matches 1. .41 of consensus"
repeat_region   30632. .31387
                /note="L2 repeat: matches 1156. .1771 of consensus"
repeat_region   31401. .31876
                /note="MLT1H repeat: matches 13. .516 of consensus"
repeat_region   31936. .32142
                /note="L2 repeat: matches 929. .1149 of consensus"
repeat_region   32790. .33087
                /note="AluSx repeat: matches 1. .297 of consensus"
misc_feature    complement(34142. .34617)
                /note="match: GSS: Em:AQ799891"
misc_feature    34145. .34705
                /note="match: GSS: Em:AQ350618
                match: STS: Em:G57679"
repeat_region   35089. .35217
                /note="L1 repeat: matches 4928. .5057 of consensus"
repeat_region   35239. .36006
                /note="L1ME1 repeat: matches 5352. .6103 of consensus"
repeat_region   37784. .37985
                /note="L1MC4 repeat: matches 5695. .5897 of consensus"
```

Query Match 4.3%; Score 56.2; DB 9; Length 169660;
Best Local Similarity 50.2%; Pred. No. 0.028;
Matches 139; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
Qy    245 TTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACAA 304
          |||   ||||||  | ||  |||   || ||  | | | |||| || | |||    | |
Db   2444 TTATATATTATATAATTATATATTATTAAATATATAAATATATAATTATATAATATATAT 2503

Qy    305 TATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGT 364
          | | |   || ||   ||||||  || ||||   ||| | | |||     |  | ||  |
Db   2504 TTTATAATTATATATTATTAAATATATAAATATATAATTATATAATTTATATTATATAAT 2563

Qy    365 AATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATA 424
           || | || |  ||  |  |    |||  || ||  ||     |  |   |  |   ||
Db   2564 TATATATTATATAATTATATATTATATAATATATAATTTATAATATATAATATATAATTA 2623

Qy    425 AACAACATAGTCTACAATGGAAAACAGTGCTTTGTACGACTATATACGCGATGTGTGGCT 484
           | || |||   |   ||  || | |   | | ||    ||||||     | | |   |
Db   2624 TATAATATATAATTTTATAAAATATAAATATGTTTAAATTTATATATAAATTATATTTAT 2683

Qy    485 ACAACATAAGACAATATAGTCGTTTGAAGATTGAACC 521
          | ||  | | | ||||||    |||  | | | || |
Db   2684 ATAATTTTATAAAATATAAATTTTTATATAATAAAAC 2720
```

```
RESULT 8
PFMAL13P6
LOCUS       PFMAL13P6             204652 bp    DNA     linear   HTG 19-AUG-1999
DEFINITION  Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL049183
VERSION     AL049183.5  GI:5763804
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 204652)
  AUTHORS   Bowman,S., Churcher,C., Harris,B.; Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Aug 24, 1999 this sequence version replaced gi:5731886.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
```

```
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            Order of segments is not known; 800 n's separate segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .204652
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="13"
BASE COUNT    68880 a  17396 c  17710 g  67856 t  32810 others
ORIGIN

  Query Match             4.2%;  Score 55.6;  DB 2;  Length 204652;
  Best Local Similarity   52.6%;  Pred. No. 0.041;
  Matches  121;  Conservative    0;  Mismatches  109;  Indels    0;  Gaps    0;

Qy     204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
           | |||   || |   || | | |  |||| | |||   | |      ||||||  | ||
Db   27538 TTCATGTGGTACAATTTGTATATATATATGTATATATATATAACAATATTATATAATTAT 27597

Qy     264 TTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATT 323
           |||||||| |  |||||||  ||    || ||||||  |  |||| ||  ||     ||||
Db   27598 ATATACATATACACACATACGTACATACATATATATAAATATATTATTACTAACAATT 27657

Qy     324 AAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACT 383
           ||||  |||||  |  | ||   ||   || |    | | ||    | | |  |
Db   27658 TAAACAAAAAAATAAAATATAATTATTCCCCAAATAAATATATGAATGTATATATATTTA 27717

Qy     384 TTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATA 433
           |  |  | | ||  || |  ||| |  |  |  | |   ||   ||||||
Db   27718 TGTCTAAATATAGTTTTTCATATATATATATATATTACTTATTAAACATA 27767


RESULT 9
AC024266/c
LOCUS       AC024266               65523 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens clone RP11-115F10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC024266
VERSION     AC024266.1  GI:7107792
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 65523)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-115F10
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65523)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L1997
                Center clone name: 115_F_10
            --------------.
            * NOTE: This record contains 74 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      793: contig of 793 bp in length
            *      794 893: gap of      100 bp
            *      894     1687: contig of 794 bp in length
            *     1688 1787: gap of      100 bp
            *     1788     2572: contig of 785 bp in length
            *     2573 2672: gap of      100 bp
            *     2673     3486: contig of 814 bp in length
            *     3487 3586: gap of      100 bp
            *     3587     4391: contig of 805 bp in length
            *     4392 4491: gap of      100 bp
            *     4492     5298: contig of 807 bp in length
            *     5299 5398: gap of      100 bp
            *     5399     6142: contig of 744 bp in length
            *     6143 6242: gap of      100 bp
            *     6243     7033: contig of 791 bp in length
            *     7034 7133: gap of      100 bp
            *     7134     7941: contig of 808 bp in length
            *     7942 8041: gap of      100 bp
            *     8042     8832: contig of 791 bp in length
            *     8833 8932: gap of      100 bp
            *     8933     9736: contig of 804 bp in length
            *     9737 9836: gap of      100 bp
            *     9837    10619: contig of 783 bp in length
            *    10620 10719: gap of      100 bp
            *    10720    11503: contig of 784 bp in length
            *    11504 11603: gap of      100 bp
            *    11604    12326: contig of 723 bp in length
            *    12327 12426: gap of      100 bp
            *    12427    13201: contig of 775 bp in length
            *    13202 13301: gap of      100 bp
            *    13302    14085: contig of 784 bp in length
            *    14086 14185: gap of      100 bp
            *    14186    14981: contig of 796 bp in length
            *    14982 15081: gap of      100 bp
            *    15082    15879: contig of 798 bp in length
            *    15880 15979: gap of      100 bp
            *    15980    16745: contig of 766 bp in length
            *    16746 16845: gap of      100 bp
            *    16846    17645: contig of 800 bp in length
            *    17646 17745: gap of      100 bp
            *    17746    18541: contig of 796 bp in length
            *    18542 18641: gap of      100 bp
            *    18642    19447: contig of 806 bp in length
```

```
*    19448 19547: gap of      100 bp
*    19548    20376: contig of 829 bp in length
*    20377 20476: gap of      100 bp
*    20477    21179: contig of 703 bp in length
*    21180 21279: gap of      100 bp
*    21280    22064: contig of 785 bp in length
*    22065 22164: gap of      100 bp
*    22165    22935: contig of 771 bp in length
*    22936 23035: gap of      100 bp
*    23036    23851: contig of 816 bp in length
*    23852 23951: gap of      100 bp
*    23952    24827: contig of 876 bp in length
*    24828 24927: gap of      100 bp
*    24928    25698: contig of 771 bp in length
*    25699 25798: gap of      100 bp
*    25799    26586: contig of 788 bp in length
*    26587 26686: gap of      100 bp
*    26687    27505: contig of 819 bp in length
*    27506 27605: gap of      100 bp
*    27606    28408: contig of 803 bp in length
*    28409 28508: gap of      100 bp
*    28509    29294: contig of 786 bp in length
*    29295 29394: gap of      100 bp
*    29395    30181: contig of 787 bp in length
*    30182 30281: gap of      100 bp
*    30282    31028: contig of 747 bp in length
*    31029 31128: gap of      100 bp
*    31129    31928: contig of 800 bp in length
*    31929 32028: gap of      100 bp
*    32029    32815: contig of 787 bp in length
*    32816 32915: gap of      100 bp
*    32916    33724: contig of 809 bp in length
*    33725 33824: gap of      100 bp
*    33825    34585: contig of 761 bp in length
*    34586 34685: gap of      100 bp
*    34686    35469: contig of 784 bp in length
*    35470 35569: gap of      100 bp
*    35570    36370: contig of 801 bp in length
*    36371 36470: gap of      100 bp
*    36471    37277: contig of 807 bp in length
*    37278 37377: gap of      100 bp
*    37378    38157: contig of 780 bp in length
*    38158 38257: gap of      100 bp
*    38258    39015: contig of 758 bp in length
*    39016 39115: gap of      100 bp
*    39116    39853: contig of 738 bp in length
*    39854 39953: gap of      100 bp
*    39954    40734: contig of 781 bp in length
*    40735 40834: gap of      100 bp
*    40835    41594: contig of 760 bp in length
*    41595 41694: gap of      100 bp
*    41695    42497: contig of 803 bp in length
*    42498 42597: gap of      100 bp
*    42598    43390: contig of 793 bp in length
*    43391 43490: gap of      100 bp
*    43491    44319: contig of 829 bp in length
*    44320 44419: gap of      100 bp
*    44420    45162: contig of 743 bp in length
*    45163 45262: gap of      100 bp
*    45263    46047: contig of 785 bp in length
*    46048 46147: gap of      100 bp
*    46148    46913: contig of 766 bp in length
*    46914 47013: gap of      100 bp
*    47014    47792: contig of 779 bp in length
*    47793 47892: gap of      100 bp
*    47893    48669: contig of 777 bp in length
*    48670 48769: gap of      100 bp
*    48770    49560: contig of 791 bp in length
*    49561 49660: gap of      100 bp
*    49661    50399: contig of 739 bp in length
*    50400 50499: gap of      100 bp
*    50500    51301: contig of 802 bp in length
*    51302 51401: gap of      100 bp
*    51402    52227: contig of 826 bp in length
*    52228 52327: gap of      100 bp
*    52328    53120: contig of 793 bp in length
*    53121 53220: gap of      100 bp
*    53221    53973: contig of 753 bp in length
*    53974 54073: gap of      100 bp
*    54074    54862: contig of 789 bp in length
*    54863 54962: gap of      100 bp
*    54963    55747: contig of 785 bp in length
*    55748 55847: gap of      100 bp
*    55848    56660: contig of 813 bp in length
*    56661 56760: gap of      100 bp
*    56761    57571: contig of 811 bp in length
*    57572 57671: gap of      100 bp
*    57672    58451: contig of 780 bp in length
*    58452 58551: gap of      100 bp
*    58552    59333: contig of 782 bp in length
*    59334 59433: gap of      100 bp
*    59434    60245: contig of 812 bp in length
*    60246 60345: gap of      100 bp
*    60346    61086: contig of 741 bp in length
*    61087 61186: gap of      100 bp
*    61187    61958: contig of 772 bp in length

  Query Match             4.2%;  Score 55.4;  DB 2;  Length 65523;
  Best Local Similarity   36.7%;  Pred. No. 0.036;
  Matches  170;  Conservative    0;  Mismatches  293;  Indels    0;  Gaps    0;

Qy     603 TAAATCCAAAGTAGGAGCGGTGCATGATGAGAATCGCTCTCAGTACTCGACATAATGAAC 662
           |
Db   21280 TNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 21221

Qy     663 CTTACGAGGTACAACAGGCAGGCAGGCAGGGACCAGGGGCCGCCTTTATTTCAGGCTCGC 722
                                                   | ||| |||| ||| |
Db   21220 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTCTTTTTTTTCTCCCTCCC 21161

Qy     723 TGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACCCCAACCTCTCACCGAAAACCG 782
             | || || ||| |  |  || ||     |  |   |||  || | | |||    |
Db   21160 CCCCGCCCCGCGCGCCCCCCCGGCGCGCCCCGGCCCCGCCCCCCCCGCCGCGCCGC 21101

Qy     783 CGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACGTCCACGGCAC 842
           ||  |    |        ||     |  ||||   ||| |   || |  || |  | |
Db   21100 GGCNCGCCGCCCGCCCGCCCCGCCCCCGCCCCCCCCCCCCCCCGGCCCCCCCCCCCCCCC 21041

Qy     843 CATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAACCACCCCCAA 902
           |  ||| | ||  ||   | ||  ||   ||  | | |    || |    ||  ||||
Db   21040 CCGCCCCCCGCCCCCCGCCCCCCCCCCCCCCCCCGCCCCCCCCCGCCCGCCCCCGCCCGC 20981

Qy     903 CCCACTCACGAAACCCCGTCCCGGCCGTGCCCGTGTCGGTCCGCGCTCGGCAACGAGGCG 962
           ||| | | |    ||||   ||  ||   |||    |   || | | |  |  | | ||
Db   20980 CCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCG 20921

Qy     963 GCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTGTCGGCCCTTTGTTTATTCATCC 1022
           |||||||  |  | |  | | |  | |||      || | ||| ||          |  ||
Db   20920 CCCCGCGCCGGCGCGCGCGCCGCCCCCCCCCGCGCCCGGCGGGGCCCCCCCCCCCNCGCCC 20861

Qy    1023 CGAAATCTCATCTGCCCCCACGGCCGACTGCGCTGCGCCGCCC 1065
           ||    | |  |  ||||  |  |||   | || || || |||
Db   20860 CGCCCCCCCCCCCCCCCCGGCCCCCGCGGGGGCGGCCCCCCCC 20818


RESULT 10
SGU36751
LOCUS       SGU36751                650 bp    DNA     linear   INV 06-FEB-1997
DEFINITION  Schizaphis graminum cytochrome oxidase subunit II (COII) gene,
            mitochondrial gene encoding mitochondrial protein, partial cds.
ACCESSION   U36751
VERSION     U36751.1  GI:1036830
KEYWORDS    .
SOURCE      greenbug.
  ORGANISM  Mitochondrion Schizaphis graminum
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha;
```

```
            Aphidiformes; Aphidoidea; Aphididae; Aphidini; Schizaphis.
REFERENCE   1  (bases 1 to 650)
  AUTHORS   Rouhbakhsh,D., Lai,C.Y., von Dohlen,C.D., Clark,M.A., Baumann,L.,
            Baumann,P., Moran,N.A. and Voegtlin,D.J.
  TITLE     The tryptophan biosynthetic pathway of aphid endosymbionts
            (Buchnera): genetics and evolution of plasmid-associated
            anthranilate synthase (trpEG) within the aphididae
  JOURNAL   J. Mol. Evol. 42 (4), 414-421 (1996)
  MEDLINE   96215866
REFERENCE   2  (bases 1 to 650)
  AUTHORS   von Dohlen,C.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-1995) Biological Sciences, Idaho State
            University, Campus Box 8007, Pocatello, ID 83209, USA
FEATURES             Location/Qualifiers
     source          1. .650
                     /organism="Schizaphis graminum"
                     /organelle="mitochondrion"
                     /db_xref="taxon:13262"
     gene            1. .650
                     /gene="COII"
     CDS             <1. .>650
                     /gene="COII"
                     /codon_start=1
                     /transl_table=5
                     /product="cytochrome oxidase subunit II"
                     /protein_id="AAB42168.1"
                     /db_xref="GI:1036831"
                     /translation="MTWLKLSFQNSNSPLMEQLIFFHDHTIFIIIMIMTMITYMMIFT
                     MKNKFINIKISENQMIEFIWTTTPPIILIFIAMPTLHLSYSMDEINCPILKIKIFGHQ
                     WFWSYEYSDFSNIEFESYMMNELNKENFRLIVVDNKTIIPFKFNIRLLISSDDVIHSW
                     TIPSLAIKIDSIPGRMNQINLFMNRPGMYFGQCSEICGINHSFMPIQIESINLNKF"
BASE COUNT      265 a     82 c     49 g    254 t
ORIGIN

  Query Match             4.2%;  Score 55.2;  DB 3;  Length 650;
  Best Local Similarity   56.4%;  Pred. No. 0.015;
  Matches  123;  Conservative    0;  Mismatches   93;  Indels    2;  Gaps    1;

Qy     235 TTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACA 294
           ||| | || ||| |||| |||| | |  |  ||     ||    |||   || || ||
Db      55 TTAATTTTTTTTCATGATCATACAATTTTTATTATTATTATAATTATAACAATAATTACT 114

Qy     295 TATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCA 354
           ||||| | |  ||||||   |   ||| ||||  ||  |||    |||| ||| | | |
Db     115 TATATAATA--ATTTTTACTATAAAAAATAAATTTATTAATATTAAAATCTCTGAAAACC 172

Qy     355 ACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATT 414
           |    |||| | |  |||   | ||||||   || ||| | | |  || |||| | |
Db     173 AAATAATTGAATTTATTTGAACTACAACTCCTCCAATTATTCTTATTTTTATTGCTATAC 232

Qy     415 TCACTCTATAAACAACATAGTCTACAATGGAAAACAGT 452
           |||||| |    |||| || | |   ||||  | |
Db     233 CTACTCTACATCTTTCATATTCAATAGATGAAATTAAT 270


RESULT 11
PFMAL13P6/c
LOCUS       PFMAL13P6             204652 bp    DNA     linear   HTG 19-AUG-1999
DEFINITION  Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL049183
VERSION     AL049183.5  GI:5763804
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 204652)
  AUTHORS   Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M.
            and Barrell,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-1999) P.falciparum Genome Sequencing Consortium,
            The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA, UK
COMMENT     On Aug 24, 1999 this sequence version replaced gi:5731886.
            For more information about this sequence or the Malaria Project,
            see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This
            sequence is unfinished and does not necessarily represent the
            correct sequence. Work on the sequence is in progress and the
            release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc.
            Order of segments is not known; 800 n's separate segments.
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .204652
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /chromosome="13"
BASE COUNT    68880 a  17396 c  17710 g  67856 t  32810 others
ORIGIN

  Query Match             4.2%;  Score 54.8;  DB 2;  Length 204652;
  Best Local Similarity   51.3%;  Pred. No. 0.063;
  Matches  153;  Conservative    0;  Mismatches  142;  Indels    3;  Gaps    1;

Qy     173 TAATGCCTTAAGCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATAT 232
           ||||     ||  ||    || ||   |   || |||     | | ||| || | ||| |
Db  154818 TAATAATAAAAATATATTATATTCTTAATAATAAATTGAACACATTTTAATTTTACATTT 154759

Qy     233 TTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCA 292
           |       | | |  |   |||| |  || |||| ||| ||  | ||| ||||    |
Db  154758 GTAATACAATATATTTTTGATATATCATAAGATATATATATATATATATATATATATA 154699

Qy     293 CATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACAC 352
           ||||||    |||| | |   |  || ||  ||  ||   ||   || |||| | | |
Db  154698 TATATATATATATATATATAATATATATATATAATGTAGTTATTTATATATTTATTATAT 154639

Qy     353 CAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAA 412
             | |||   ||   ||||| |   | |     |||   | |||| | ||||   ||
Db  154638 ATTATATATTTGTATAGATTCTCAATTTAATGCCTATATGAAAAATTGATATATT---AA 154582

Qy     413 TTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGTACGACTATATA 470
           ||||  | || |||||| |||    | ||  | ||  | |  | |   |    || |||
Db  154581 TTTCCATATAAAAACAATATATATAAAAAAAGTAATAAATAATATAAATATATAAATA 154524


RESULT 12
ATT25K17
LOCUS       ATT25K17              89904 bp    DNA     linear   PLN 27-AUG-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 4, BAC clone  (ESSA project).
ACCESSION   AL049171
VERSION     AL049171.1  GI:4539415
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 89904)
  AUTHORS   Bevan,M., Koetter,P., Hempel,S., Entian,K.-D., Bancroft,I.,
            Mewes,H.W., Mayer,K.F.X. and Schueller,C.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89904)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            schuelle@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
```

```
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosome 4 can be
            viewed at: http://websvr.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .89904
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     misc_feature    1. .11070
                     /note="overlap to BAC ; please refer to this entry for
                     analysis and annotation"
     exon            19417. .19605
                     /gene="T25K17.10"
                     /number=1
     gene            19417. .21056
                     /gene="T25K17.10"
     CDS             join(19417. .19605,19706. .20001,20198. .21056)
                     /gene="T25K17.10"
                     /EC_number="4.4.1.14"
                     /note="strong similarity to ACC synthase , Malus
                     domestica, U73816
                     Contains Aminotransferases class-I pyridoxal-phosphate
                     attachmentsite, Aa_Transfer_Class_1 [SLSKDLGLPGFRVG]
                     contains EST gb:N65643, AA650853, Z33841"
                     /codon_start=1
                     /product="1-aminocyclopropane-1-carboxylate synthase-like
                     protein"
                     /protein_id="CAB38949.1"
                     /db_xref="GI:4539416"
                     /translation="MGLPLMMERSSNNNNVELSRVAVSDTHGEDSPYFAGWKAYDENP
                     YDESHNPSGVIQMGLAENQVSFDLLETYLEKKNPEGSMWGSKGAPGFRENALFQDYHG
                     LKTFRQAMASFMEQIRGGKARFDPDRIVLTAGATAANELLTFILADPNDALLVPTPYY
                     PGFDRDLRWRTGVKIVPIHCDSSNHFQITPEALESAYQTARDANIRVRGVLITNPSNP
                     LGATVQKKVLEDLLDFCVRKNIHLVSDEIYSGSVFHASEFTSVAEIVENIDDVSVKER
                     VHIVYSLSKDLGLPGFRVGTIYSYNDNVVRTARRMSSFTLVSSQTQHMLASMLSDEEF
                     TEKYIRINRERLRRRYDTIVEGLKKAGIECLKGNAGLFCWMNLGFLLEKKTKDGELQL
                     WDVILKELNLNISPGSSCHCSEVGWFRVCFANMSENTLEIALKRIHEFMDRRRRF"
     intron          19606. .19705
                     /gene="T25K17.10"
                     /number=1
     exon            19706. .20001
                     /gene="T25K17.10"
                     /number=2
     intron          20002. .20197
                     /gene="T25K17.10"
                     /number=2
     exon            20198. .21056
                     /gene="T25K17.10"
                     /number=3
     gene            26480. .27228
                     /gene="T25K17.20"
     CDS             join(26480. .26661,27042. .27228)
                     /gene="T25K17.20"
                     /note="strong similarity to predicted protein, Arabidopsis
                     thaliana, AC005169"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB38950.1"
                     /db_xref="GI:4539417"
                     /translation="MASKLLQLKSKACEASKFVSKHGTTYYKQLLDKNKMYIQEPATI
                     EKCNELSKQLLYTRLASIPGRSESFWKEVNHVKGLWKNRADLKVEDAGIAALFGLECF
                     AWYCAGEIVGRGFTFTGYYP"
     exon            26480. .26661
                     /gene="T25K17.20"
                     /number=1
     intron          26662. .27041
                     /gene="T25K17.20"
                     /number=1
     exon            27042. .27228
                     /gene="T25K17.20"
                     /number=2
     exon            28289. .28369
                     /gene="T25K17.30"
                     /number=1
     gene            28289. .29256
                     /gene="T25K17.30"
     CDS             join(28289. .28369,28417. .28641,28748. .28879,28966. .29256)
                     /gene="T25K17.30"
                     /note="strong similarity to caffeoyl-CoA
                     O-methyltransferase, Pinus taeda, AF036095"
                     /codon_start=1
                     /product="caffeoyl-CoA O-methyltransferase-like protein"
                     /protein_id="CAB38951.1"
                     /db_xref="GI:4539418"
                     /translation="MAKDEAKGLLKSEELYKVCPYILISTLYILETSVYPREPEVLRE
                     LRNITHNHPQAGMATAPDAGQLMGMLLNLVNARKTIEVGVFTGYSLLLTALTLPEDGK
                     VIAIDMNRDSYEIGLPVIKKAGVEHKIDFKESEALPALDELLNNKVNEGGFDFAFVDA
                     DKLNYWNYHERLIRLIKVGGIIVYDNTLWGGSVAEPDSSTPEWRIEVKKATLELNKKL
                     SADQRVQISQAALGDGITICRRLY"
     intron          28370. .28416
                     /gene="T25K17.30"
                     /number=1
     exon            28417. .28641
                     /gene="T25K17.30"
                     /number=2
     intron          28642. .28747
                     /gene="T25K17.30"
                     /number=2
     exon            28748. .28879
                     /gene="T25K17.30"
                     /number=3
     intron          28880. .28965
                     /gene="T25K17.30"
                     /number=3
     exon            28966. .29256
                     /gene="T25K17.30"
                     /number=4
     tRNA            29533. .29604
                     /note="Pro-tRNA_(TGG)"
                     /product="tRNA-Pro"
     exon            30136. .30215
                     /gene="T25K17.40"
                     /number=1
     gene            30136. .30663
                     /gene="T25K17.40"
     CDS             join(30136. .30215,30384. .30663)
                     /gene="T25K17.40"
                     /note="strong similarity to ribosomal protein L31,
                     Nicotiana glutinosa, U23784
                     contains EST gb:H77007"
                     /codon_start=1
                     /product="putative ribosomal protein"
                     /protein_id="CAB38952.1"
                     /db_xref="GI:4539419"
                     /translation="MSEKKGRKEEVVTREYTINLHRRLHSCTFKKKAPKAIKEIRKFA
                     EKEMGTKDVRVDVKLNKQIWSKGIRGPPRRIRVRVARKRNDDEDAKEEFFSLVTVAEI
                     PAEGLSGLGTKIIEEED"
     intron          30216. .30383
                     /gene="T25K17.40"
                     /number=1
     exon            30384. .30663
                     /gene="T25K17.40"
                     /number=2
     misc_feature    30910. .31193
                     /note="EST AA586104 matches to position 30910-31193; no
                     obvious genmodel"
     gene            32558. .33006
                     /gene="T25K17.50"
     exon            complement(32558. .32659)
                     /gene="T25K17.50"
                     /number=1
     gene            complement(join(32558. .32659,32745. .32806,32898. .33006))
                     /gene="T25K17.50"
     CDS             complement(join(<32558. .32659,32745. .32806,32898. .33006))
```

```
                     /gene="T25K17.50"
                     /note="contains EST gb:F45349"
                     /codon_start=1
                     /product="putative protein (fragment)"
                     /protein_id="CAB38953.1"
                     /db_xref="GI:4539420"
                     /translation="MKKLARKWRRTRRDEDDDKFVLPTSEDVDSRPIDTQEQEEYVRY
                     LEEAHAQQSLQWRSVFVVLLICFGAFLFYSIFQQFMSPWELVLIISP"
     intron          complement(32660. .32744)
                     /gene="T25K17.50"
                     /number=1
     exon            complement(32745. .32806)
                     /gene="T25K17.50"
                     /number=2
     intron          complement(32807. .32897)
                     /gene="T25K17.50"
                     /number=2
     exon            complement(32898. .33006)
                     /gene="T25K17.50"
                     /number=3
     gene            33751. .35163
                     /gene="T25K17.60"
     CDS             join(33751. .34158,34325. .34648,34729. .34863,34945. .35163)
                     /gene="T25K17.60"
                     /note="strong similarity to WSI76 protein induced by water
                     stress, Oryzasativa, D26537"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB38954.1"
                     /db_xref="GI:4539421"
                     /translation="MAQMSMTVEKSIKADVTVSHDRVKRAYVTFLAGNKDYWMGVVGL
                     AKGLRKVKSAYPLVVAILPDVPEEHRQILLAQGCIIREIEPVYPPENKTGYSMAYYVI

  Query Match             4.1%;  Score 54;  DB 8;  Length 89904;
  Best Local Similarity   56.6%;  Pred. No. 0.082;
  Matches  120;  Conservative    0;  Mismatches   90;  Indels    2;  Gaps    1;

Qy     259 AGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATT 318
           | || |  |  ||| |   |     | || ||||  ||| | |   |  |||  ||  ||
Db   25248 ATTAATATTTAATATAATAATTACTAAATAATCATTTATTTGAATTTTGTTTATTATTTT 25307

Qy     319 AAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAA 378
            | |||||| |||  |   || |  | ||| |  |   | ||| |||  |||||| |||
Db   25308 CATTTAAAATTAAAGTTCTGTATAAATTTAAAATAATTTATATTATAAATTTTTCTTCAA 25367

Qy     379 CAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTA 438
           |  ||||  |||| || |  ||   ||||  ||||  | |  | |   || |||   |
Db   25368 AACATTTATTTATTTTATA--GTAAAATTTTCAATTATATTACAAAGTTAAAATAAGTTT 25425

Qy     439 CAATGGAAAACAGTGCTTTGTACGACTATATA 470
           || |  |||||| |  ||| ||| | | || |
Db   25426 CACTTTAAAACATTTATTTATACTAATTTAAA 25457


RESULT 13
ATCHRIV64
LOCUS       ATCHRIV64              196286 bp    DNA     linear   PLN 16-MAR-2000
DEFINITION  Arabidopsis thaliana DNA chromosome 4, contig fragment No. 64.
ACCESSION   AL161564
VERSION     AL161564.2  GI:7269427
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 50686 to 155423; 144355 to 196286)
  AUTHORS   Rose,M., Hempel,S., Entian,K.-D., Mewes,H.W., Lemcke,K. and
            Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   2  (bases 77316 to 77804)
  AUTHORS   Volckaert,G., Grymonprez,B., Voet,M., Robben,J., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 82307)
  AUTHORS   Peters,S.A., van Staveren,M., Dirkse,W., Stiekema,W., Mewes,H.W.,
            Lemcke,K. and Mayer,K.F.X.
  JOURNAL   Unpublished
REFERENCE   4  (bases 1 to 196286)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
            Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
            E-mail: michael.bevan@bbsrc.ac.uk
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
            this fragment has an overlap with ATCHRIV63 at the 5' end and an
            overlap with ATCHRIV65 at the 3' end.
FEATURES             Location/Qualifiers
     source          1. .196286
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="4"
     gene            6687. .8444
                     /gene="AT4g25770"
     exon            complement(6687. .6833)
                     /gene="AT4g25770"
                     /number=1
     gene            complement(join(6687. .6833,6920. .7036,7083. .7163,
                     7278. .7370,7494. .7733,7782. .7882,7945. .8042,8134. .8444))
                     /gene="AT4g25770"
     CDS             complement(join(6687. .6833,6920. .7036,7083. .7163,
                     7278. .7370,7494. .7733,7782. .7882,7945. .8042,8134. .8444))
                     /gene="AT4g25770"
                     /note="similarity to hypothetical protein T27I1.6 -
                     Arabidopsis thaliana, PIR2:T00623"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB79432.1"
                     /db_xref="GI:7269428"
                     /translation="MIASFSRFQISIPDLIPSSPRFLPPSQNQFRFRLYLQFPAQPKN
                     LCQMEWMNRLGIGCFAANRSKRELTTEIDNGGEDFFDADVMESAEKPDHLVVMVNGIV
                     GSAADWKYAAEQFVKKFPDKVLVHRACFLTYRYCGSESNSATLTFDGVDKMGERLANE
                     VVFGCIVVSSVLGVVKHRSGLKKISFVAHSLGGLVARYAIGKLYEQPGEVDSLDSPSK
                     EKSARGGEIAGLEPMNFITFATPHLGSRGHRQFPILCGLPFLERTASQTAHLAAGRTG
                     KHLFLIDMLIISSYDLNRSALNAFKRRVAYANRLSMVFDYFYHLQEFALTNEKAMVGW
                     RTSSIRRPNELPKPNLLATDPNYPHIVYVERGNVDNGSCQSTSTVVTEQDTDLEGFAL
                     FLT"
     intron          complement(6834. .6919)
                     /gene="AT4g25770"
                     /number=1
     exon            complement(6920. .7036)
                     /gene="AT4g25770"
                     /number=2
     intron          complement(7037. .7082)
                     /gene="AT4g25770"
                     /number=2
     exon            complement(7083. .7163)
                     /gene="AT4g25770"
                     /number=3
     intron          complement(7164. .7277)
                     /gene="AT4g25770"
                     /number=3
     exon            complement(7278. .7370)
                     /gene="AT4g25770"
                     /number=4
     intron          complement(7371. .7493)
                     /gene="AT4g25770"
                     /number=4
     exon            complement(7494. .7733)
                     /gene="AT4g25770"
```

```
                /number=5
intron          complement(7734. .7781)
                /gene="AT4g25770"
                /number=5
exon            complement(7782. .7882)
                /gene="AT4g25770"
                /number=6
intron          complement(7883. .7944)
                /gene="AT4g25770"
                /number=6
exon            complement(7945. .8042)
                /gene="AT4g25770"
                /number=7
intron          complement(8043. .8133)
                /gene="AT4g25770"
                /number=7
exon            complement(8134. .8444)
                /gene="AT4g25770"
                /number=8
CDS             9569. .10141
                /gene="AT4g25780"
                /note="strong similarity to gene PR-1 protein - Medicago
                truncatula, PIR2:S47171
                Contains Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7
                signatures AA147-157;Extracellular proteins
                SCP/Tpx-1/Ag5/PR-1/Sc7 signatures AA173-184
                contains EST gb:Z26548, AA586201, AA712838, AA712715"
                /codon_start=1
                /product="putative pathogenesis-related protein"
                /protein_id="CAB79433.1"
                /db_xref="GI:7269429"
                /translation="MSSSYLRVILLLGALNVVVSLSITNSLITKSATLGQVFRICKNL
                CPGCDHDSLQFLFRHNLVRAARFEPPLIWDRRLQNYAQGWANQRRGDCALRHSVSNGE
                FNLGENIYWGYGANWSPADAVVAWASEKRFYHYGSNTCDAGQMCGHYTQIVWKSTRRV
                GCARVVCDNGGIFMTCNYDPPGNYIGQKPY"
exon            9569. .10141
                /gene="AT4g25780"
                /number=1
gene            9569. .10141
                /gene="AT4g25780"
exon            complement(10710. .11074)
                /gene="AT4g25790"
                /number=1
gene            10710. .11752
                /gene="AT4g25790"
gene            complement(join(10710. .11074,11485. .11752))
                /gene="AT4g25790"
CDS             complement(join(10710. .11074,11485. .11752))
                /gene="AT4g25790"
                /note="strong similarity to gene PR-1 protein - Medicago
                truncatula, PIR2:S47171
                Contains Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7
                signatures AA166-176;Extracellular proteins
                SCP/Tpx-1/Ag5/PR-1/Sc7 signatures AA193-204
                contains EST gb:AA712715, N65358"
                /codon_start=1
                /product="putative pathogenesis-related protein"
                /protein_id="CAB79434.1"
                /db_xref="GI:7269430"
                /translation="MATTTSTQAIVFAIALFLVAIQAVHADYYRHRPQRPITPSPPPY
                VAKPQPLPTPSPKPILYQPPPTYQPPTGSFEQQFLDPHNTVRGGLGLPPLVWDVKIAS
                YATWWANQRRYDCSLTHSTGPYGENLFWGSGSDFTSTFAVESWTVEAKSYNHMTNTCE
                GDGMCGHYTQIVWRETRRLGCARVVCENGAGVFITCNYDPPGNYVGEKPY"
intron          complement(11075. .11484)
                /gene="AT4g25790"
                /number=1
exon            complement(11485. .11752)
                /gene="AT4g25790"
                /number=2
gene            13985. .16227
                /gene="AT4g25800"
CDS             join(13985. .14010,14239. .14669,14761. .14921,15016. .15219,
                15390. .15595,15645. .16227)
                /gene="AT4g25800"
                /note="EST GB:F14385 in 5' untranslated region
                strong similarity to calmodulin-binding protein (TCB60) -
                Nicotianatabacum, PID:g1698548
                contains EST gb:F14386, F14385, N96168, N96141"
                /codon_start=1
                /product="putative calmodulin-binding protein"
                /protein_id="CAB79435.1"
                /db_xref="GI:7269431"
                /translation="MNQFCLLVRLSLPLFTGGRVEGEQGATIHVVLIDANTGRPVTVG
                PEASLKLEVVVLGGDFNNEDDEDWTQEEFESHVVKEREGKRPLLTGDLFVVLKEGVGT
                LGEIVFTDNSSWIRSRKFRLGLRVPSGYCDGIRIREAKTEAFSVKDHRGELYKKHYPP
                ALNDEVWRLEKIGKDGAFHKRLTAAGIVTVEGFLRQLVRDSTKLRAILGSGMSNKMWD
                LLVEHAKTCVLSGKLYIYYTEDSRSVGVVFNNIYELSGLITEDQYLSADSLSESQKVY
                VDGLVKKAYENWNQVVEYEGESLLNLNQPERLDISQTDPVTALASYSTVPLSQFPEFA
                IEGWQISPFLYSTSIFTAGYNQTLTTALPHNPQAQFDFVPQQDQFIGIQQPQTQTNIE
                NENVTRLVLGPPQSSTGGYQDIKSSADQENLNPFEDWTNLSENDFFSEEEIRQTSHDL
                LANEDMQQLLFSMGGGKGEDGFTFPSFMQNTPMMQGYDEEGRGRSGKAVVGWLKVKAA
                MRWGFFIRRKAAERRAQIVELHDNNDGK"
exon            13985. .14010
                /gene="AT4g25800"
                /number=1
intron          14011. .14238
                /gene="AT4g25800"
                /number=1
exon            14239. .14669
                /gene="AT4g25800"
                /number=2
intron          14670. .14760
                /gene="AT4g25800"
                /number=2
exon            14761. .14921
                /gene="AT4g25800"
                /number=3
intron          14922. .15015
                /gene="AT4g25800"
                /number=3
```

```
  Query Match             4.1%;  Score 54;  DB 8;  Length 196286;
  Best Local Similarity   56.6%;  Pred. No. 0.097;
  Matches  120;  Conservative    0;  Mismatches   90;  Indels    2;  Gaps    1;

Qy     259 AGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATT 318
           | || |  |  |||  |    |      | || ||||  ||| | |   |  |||  ||  ||
Db  169602 ATTAATATTTAATATAATAATTACTAAATAATCATTTATTTGAATTTTGTTTATTATTTT 169661

Qy     319 AAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAA 378
            | |||||| ||||  |   || |  | ||| |  |   | ||| |||  |||||| |||
Db  169662 CATTTAAAATTAAAGTTCTGTATAAATTTAAAATAATTTATATTATAAATTTTTCTTCAA 169721

Qy     379 CAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTA 438
            |  ||||  |||| || |  ||   |||| ||||  | |  | |   || |||   |
Db  169722 AACATTTATTTATTTTATA--GTAAAATTTTCAATTATATTACAAAGTTAAAATAAGTTT 169779

Qy     439 CAATGGAAAACAGTGCTTTGTACGACTATATA 470
           || |  |||||| |  ||| ||| | | || |
Db  169780 CACTTTAAAACATTTATTTATACTAATTTAAA 169811


RESULT 14
AC006248/c
LOCUS       AC006248              115522 bp    DNA     linear   PLN 05-APR-2000
DEFINITION  Arabidopsis thaliana chromosome II section 91 of 255 of the
            complete sequence. Sequence from clones F26H6, F9O13, F19G14.
ACCESSION   AC006248 AE002093
VERSION     AC006248.3  GI:6598542
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 115522)
```

AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M., Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L., Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.

TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana

JOURNAL Nature 402 (6763), 761-768 (1999)

MEDLINE 20083487

PUBMED 10617197

REFERENCE 2 (bases 1 to 115522)

AUTHORS Lin,X.

TITLE Direct Submission

JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

COMMENT On Dec 17, 1999 this sequence version replaced gi:4335711.
The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES Location/Qualifiers

source 1. .115522
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"

misc_feature complement(<1. .1408)
/note="Sequence from clone F26H6"

repeat_region 515. .628
/rpt_family="(TAAAA)n"

mRNA join(<940. .1083,1286. .1597,1804. .2328,2426. .>2653)
/gene="At2g15460"

gene <940. .>2653
/gene="At2g15460"
/note="F26H6.2"

CDS join(940. .1083,1286. .1597,1804. .2328,2426. .2653)
/gene="At2g15460"
/codon_start=1
/product="putative polygalacturonase"
/protein_id="AAD17390.2"
/db_xref="GI:6598543"
/translation="MACISFAFKALCFCLFFIVVASRSTVRPKVFNVQRHGAKPDGKT
DNANAFTSIWSRACKRISGSSKIYVPKGTFYLGGVEFVGPCKNPIEFIIDGTLLAPAN
PSDIKQDTWINFRYINNLSISGSGTLDGQGKQSWPHNDCHTNPNCPKLAMTMGFAFVN
NSNIKDITSLNSKMGHFNFFSVHHFNITGVTITAPGDSPNTDGIKMGSCSNIQISDTN
IGTGDDCIAILSGTTNLNISNVNCGPGHGISVGSLGKSKDEKDVKDLIVRDVIFNGTS
DGIRIKTWESSASKILVSNFVYENIQMIDVGKPINIDQKYCPHPPCEHERKGESHVQI
QNLKLKNIYGTSKNKVAVNLQCSKIFPCKNVELIDINIKQNGVKDGSSTSVCENVDGF
ARGKMFPRIV"

misc_feature 1409. .115321
/note="Sequence from clone F9O13"

repeat_region complement(2357. .2378)
/rpt_family="(TAAAA)n"

repeat_region 3911. .4024
/rpt_family="(TAAAA)n"

mRNA join(<4336. .4479,4682. .4993,5202. .5726,5830. .>6057)
/gene="At2g15470"

gene <4336. .>6057
/gene="At2g15470"
/note="F9O13.2"

CDS join(4336. .4479,4682. .4993,5202. .5726,5830. .6057)
/gene="At2g15470"
/codon_start=1
/product="putative polygalacturonase"
/protein_id="AAD17391.1"
/db_xref="GI:4335713"
/translation="MACISFAFKALCFCLFFIVVASRSTVRPKVFNVQRHGAKPDGKT
DNANAFTSIWSRACKRISGSSKIYVPKGTFYLGGVEFVGLCKNPIEFIIDGTLLAPAN
PSDIKQDTWINFRYINNLSISGSGTLDGQGKQSWPHNDCHTNPNCPKLAMTMGFAFVN
NSNIKDITSLNSKMGHFNFFSVHHFNITGVTITAPGDSPNTDGIKMGSCSNIQISDTN
IGTGDDCIAILSGTTNLNISNVNCGPGHGISVGSLGKNKDEKDAKDLIVRDVIFNGTS
DGIRIKTWESSASKILVSNFVYENIQMIDVGKPINIDQKYCPHPPCEHERKSHVQIQN
LKLKNIYGTSKNKVAMNLQCSKIFPCKNVELIDINIKQNGVKDGSSTSVCENVDGFAR
GKMFPQHCLN"

repeat_region complement(5755. .5776)
/rpt_family="(TAAAA)n"

repeat_region 7220. .7281
/rpt_family="(TA)n"

repeat_region complement(7698. .7802)
/rpt_family="(TA)n"

repeat_region complement(7735. .7838)
/rpt_family="(TAAA)n"

repeat_region complement(8026. .8142)
/rpt_family="(TAA)n"

repeat_region complement(8080. .8183)
/rpt_family="(GAAAA)n"

repeat_region complement(8739. .8780)
/rpt_family="(CAAAA)n"

mRNA join(<8823. .9813,9995. .>10458)
/gene="At2g15480"

gene <8823. .>10458
/gene="At2g15480"
/note="F9O13.3"

CDS join(8823. .9813,9995. .10458)
/gene="At2g15480"
/codon_start=1
/product="putative glucosyltransferase"
/protein_id="AAD17392.1"
/db_xref="GI:4335714"
/translation="MNREVSERIHILFFPFMAQGHMIPILDMAKLFSRRGAKSTLLTT
PINAKIFEKPIEAFKNQNPDLEIGIKIFNFPCVELGLPEGCENADFINSYQKSDSGDL
FLKFLFSTKYMKQQLESFIETTKPSALVADMFFPWATESAEKLGVPRLVFHGTSFFSL
CCSYNMRIHKPHKKVATSSTPFVIPGLPGDIVITEDQANVAKEETPMGKFMKEVRESE
TNSFGVLVNSFYELESAYADFYRSFVAKRAWHIGPLSLSNRELGEKARRGKKANIDEQ
ECLKWLDSKTPGSVVYLSFGSGTNFTNDQLLEIAFGLEGSGQSFIWVVRKNENQGDNE
EWLPEGFKERTTGKGLIIPGWAPQVLILDHKAIGGFVTHCGWNSAIEGIAAGLPMVTW
PMGAEQFYNEKLLTKVLRIGVNVGATELVKKGKLISRAQVEKAVREVIGGEKAEERRL
WAKKLGEMAKAAVEEGGSSYNDVNKFMEELNGRK"

```
     mRNA            join(<11828. .12737,12932. .>13404)
                     /gene="At2g15490"
     gene            <11828. .>13404
                     /gene="At2g15490"
                     /note="F9O13.4"
     CDS             join(11828. .12737,12932. .13404)
                     /gene="At2g15490"
                     /codon_start=1
                     /product="putative glucosyltransferase"
                     /protein_id="AAD17393.1"
                     /db_xref="GI:4335715"
                     /translation="MAKLFARRGAKSTLLTTPINAKILEKPIEAFKVQNPDLEIGIKI
                     LNFPCVELGLPEGCENRDFINSYQKSDSFDLFLKFLFSTKYMKQQLESFIETTKPSAL
                     VADMFFPWATESAEKIGVPRLVFHGTSSFALCCSYNMRIHKPHKKVASSSTPFVIPGL
                     PGDIVITEDQANVTNEETPFGKFWKEVRESETSSFGVLVNSFYELESSYADFYRSFVA
                     KKAWHIGPLSLSNRGIAEKAGRGKKANIDEQECLKWLDSKTPGSVVYLSFGSGTGLPN
                     EQLLEIAFGLEGSGQNFIWVVSKNENQVGTGENEDWLPKGFEERNKGKGLIIRGWAPQ
                     VLILDHKAIGGFVTHCGWNSTLEGIAAGLPMVTWPMGAEQFYNEKLLTKVLRIGVNVG
                     ATELVKKGKLISRAQVEKAVREVIGGEKAEERRLRAKELGEMAKAAVEEGGSSYNDVN
                     KFMEELNGRK"
     mRNA            join(<13713. .13796,13945. .14093,14268. .14303,14375. .14413,
                     14667. .>14757)
                     /gene="At2g15500"
     gene            <13713. .>14757
                     /gene="At2g15500"
                     /note="F9O13.5"
     CDS             join(13713. .13796,13945. .14093,14268. .14303,14375. .14413,
                     14667. .14757)
                     /gene="At2g15500"
                     /codon_start=1
                     /product="putative poly(A)-binding protein"
                     /protein_id="AAD17394.1"
                     /db_xref="GI:4335716"
                     /translation="MSKLIGMLLNDKQVYVGPFLRGKKGILLAVESLKGNTFDDKQWV
                     VDRAQKKSEREMELKVPYEQSLKEIADKFQSYNFRGSGFVAFSTQNVTISGICKDVID
                     ESVAMPPSVGPRMPIYPLGSPGIRKSMFYG"
     mRNA            complement(join(<16628. .18427,18538. .18661,18914. .19100,
                     19343. .19660,19695. .>20682))
                     /gene="At2g15510"
     gene            complement(<16628. .>20682)
                     /gene="At2g15510"
                     /note="F9O13.6"

  Query Match             4.1%;  Score 53.8;  DB 8;  Length 115522;
  Best Local Similarity   53.3%;  Pred. No. 0.096;
  Matches  136;  Conservative    0;  Mismatches  117;  Indels    2;  Gaps    1;

Qy     226 TGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACAT 285
           ||  |||| | || ||   |  |      |  |  | ||||  |  | | |   |     |
Db    8079 TGATTATTGTAATAGTAATCAAAAAGGTGAATTAATAGTATTTAACATAATAATTAGTAA 8020

Qy     286 ATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTT 345
           |  ||||  ||| | |   |  |||  ||| || | |||||| ||||  |     | |  |
Db    8019 AAAATCATTTATTTGAATTTTGTTTATTAATTTCATTTAAAATTAAAGTTCTGCATAAAT 7960

Qy     346 CTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTA 405
           |||  |  ||  | ||| |||  || ||| ||| |  ||||  |||| || | |     |
Db    7959 TTAAAATAAATTATATTATAAATTTCTCTTCAAAACATTTATTTATTTTATAATAAAAAA 7900

Qy     406 TTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGTACGACT 465
           ||   ||||  | |  | |   || |||   | || |  ||||||  |  ||| ||| | |
Db    7899 TT--AAATTATATTACAAATTTAAAATAAGTTTCACTTTAAAACATTTATTTATACTAAT 7842

Qy     466 ATATACGCGATGTGT 480
           || |    || | |
Db    7841 TTAAAATAAATTTAT 7827


RESULT 15
AC105576/c
LOCUS       AC105576               214033 bp    DNA     linear   HTG 09-JAN-2002
DEFINITION  Rattus norvegicus clone CH230-236F15, WORKING DRAFT SEQUENCE, 50
            unordered pieces.
ACCESSION   AC105576
VERSION     AC105576.1  GI:18092798
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 214033)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 214033)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     ---------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            ---------- Project Information
                Center project name: GNFN
                Center clone name: CH230-236F15
            ---------- Summary Statistics
                Assembly program: Phrap; version 0.990329First call to
            findPhrapList
                Consensus quality: 185044 bases at least Q40
                Consensus quality: 190296 bases at least Q30
                Consensus quality: 194221 bases at least Q20
                Estimated insert size: 190368; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 3.5x in Q20 bases; sum-of-contigs estimation
```

```
--------------
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 50 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1    16332: contig of 16332 bp in length
*   16333    16432: gap of unknown length
*   16433    26891: contig of 10459 bp in length
*   26892    26991: gap of unknown length
*   26992    39114: contig of 12123 bp in length
*   39115    39214: gap of unknown length
*   39215    47706: contig of 8492 bp in length
*   47707    47806: gap of unknown length
*   47807    54905: contig of 7099 bp in length
*   54906    55005: gap of unknown length
*   55006    61730: contig of 6725 bp in length
*   61731    61830: gap of unknown length
*   61831    68922: contig of 7092 bp in length
*   68923    69022: gap of unknown length
*   69023    73955: contig of 4933 bp in length
*   73956    74055: gap of unknown length
*   74056    79739: contig of 5684 bp in length
*   79740    79839: gap of unknown length
*   79840    87203: contig of 7364 bp in length
*   87204    87303: gap of unknown length
*   87304    95716: contig of 8413 bp in length
*   95717    95816: gap of unknown length
*   95817   101220: contig of 5404 bp in length
*  101221   101320: gap of unknown length
*  101321   106991: contig of 5671 bp in length
*  106992   107091: gap of unknown length
*  107092   111184: contig of 4093 bp in length
*  111185   111284: gap of unknown length
*  111285   117002: contig of 5718 bp in length
*  117003   117102: gap of unknown length
*  117103   122719: contig of 5617 bp in length
*  122720   122819: gap of unknown length
*  122820   127676: contig of 4857 bp in length
*  127677   127776: gap of unknown length
*  127777   132248: contig of 4472 bp in length
*  132249   132348: gap of unknown length
*  132349   137419: contig of 5071 bp in length
*  137420   137519: gap of unknown length
*  137520   140519: contig of 3000 bp in length
*  140520   140619: gap of unknown length
*  140620   145613: contig of 4994 bp in length
*  145614   145713: gap of unknown length
*  145714   150743: contig of 5030 bp in length
*  150744   150843: gap of unknown length
*  150844   153229: contig of 2386 bp in length
*  153230   153329: gap of unknown length
*  153330   155705: contig of 2376 bp in length
*  155706   155805: gap of unknown length
*  155806   159641: contig of 3836 bp in length
*  159642   159741: gap of unknown length
*  159742   161937: contig of 2196 bp in length
*  161938   162037: gap of unknown length
*  162038   166136: contig of 4099 bp in length
*  166137   166236: gap of unknown length
*  166237   169654: contig of 3418 bp in length
*  169655   169754: gap of unknown length
*  169755   173209: contig of 3455 bp in length
*  173210   173309: gap of unknown length
*  173310   174872: contig of 1563 bp in length
*  174873   174972: gap of unknown length
*  174973   178370: contig of 3398 bp in length
*  178371   178470: gap of unknown length
*  178471   181766: contig of 3296 bp in length
*  181767   181866: gap of unknown length
*  181867   184177: contig of 2311 bp in length
*  184178   184277: gap of unknown length
*  184278   186253: contig of 1976 bp in length
*  186254   186353: gap of unknown length
*  186354   188889: contig of 2536 bp in length
*  188890   188989: gap of unknown length
*  188990   190730: contig of 1741 bp in length
*  190731   190830: gap of unknown length
*  190831   192237: contig of 1407 bp in length
*  192238   192337: gap of unknown length
*  192338   194466: contig of 2129 bp in length
*  194467   194566: gap of unknown length
*  194567   196224: contig of 1658 bp in length
*  196225   196324: gap of unknown length
*  196325   197794: contig of 1470 bp in length
*  197795   197894: gap of unknown length
*  197895   199371: contig of 1477 bp in length
*  199372   199471: gap of unknown length
*  199472   202264: contig of 2793 bp in length
*  202265   202364: gap of unknown length
*  202365   204373: contig of 2009 bp in length
*  204374   204473: gap of unknown length
*  204474   205726: contig of 1253 bp in length
*  205727   205826: gap of unknown length
*  205827   207670: contig of 1844 bp in length
*  207671   207770: gap of unknown length
*  207771   208820: contig of 1050 bp in length
*  208821   208920: gap of unknown length
*  208921   210151: contig of 1231 bp in length
*  210152   210251: gap of unknown length
*  210252   211652: contig of 1401 bp in length
*  211653   211752: gap of unknown length
*  211753   212869: contig of 1117 bp in length
*  212870   212969: gap of unknown length
*  212970   214033: contig of 1064 bp in length.
FEATURES             Location/Qualifiers
     source          1. .214033
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="CH230-236F15"
BASE COUNT    63123 a  41277 c  42183 g  62289 t   5161 others
ORIGIN

  Query Match             4.1%;  Score 53.6;  DB 2;  Length 214033;
  Best Local Similarity   50.8%;  Pred. No. 0.12;
  Matches  128;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy    814 CCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCC 873
          |||  |  | | |  | || ||    | || | | ||| | | | | | ||| ||   ||
Db 140327 CCCCCGAACCCCCGTCCCGCCCCGCCCCCCCTTCGTCTCCTGACCCCCCCCCACCCCCCCC 140268

Qy    874 GTGTCGCGGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCC 933
            | |    |  | |  ||  |||   ||| | | |     |||||||  |||   ||
Db 140267 TCCCCTCCCCCCCCCCCCTCCCGCCCGCGCCCCCCCGCCCTCGCCCGTCCTCGCCCCCCC 140208

Qy    934 CGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGA 993
          |  |   |  | ||| |  |  || | |  ||  ||  |||| | || |    | |||
Db 140207 CCCGCTCGCTCCCGCCCCCCCCCGCGCCCTCCGTCGGGGCTGCGGCCGCCCCCCGCCCCG 140148

Qy    994 CACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCACGGCCGACTGC 1053
            ||||  || ||| | |    |||  |||    || |     || || ||  |  || |
Db 140147 TCCCCTCCCGCCCCCTCGCGCCTTCCGCCCCCCCTCCCGCTCCCCTCCCCGCTCCCCTCC 140088

Qy   1054 GCTGCGCCGCCC 1065
           | ||||| |||
Db 140087 CCCGCGCCCCCC 140076


Search completed: November  7, 2002, 12:45:06
Job time : 3522.86 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:29:53 ; Search time 166.436 Seconds
(without alignments)
13503.300 Million cell updates/sec

Title: US-09-905-558C-16
Perfect score: 1309
Sequence: 1 cccatcgctgctttgtctac.....caagggaaagtgatcccatg 1309

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	235	18.0	12313	21	AAZ35184	Corn delta-12 desa
	2	232.8	17.8	2126	24	AAS96571	Corn promoter sequ
c	3	124.6	9.5	1888	22	AAF81468	Corn promoter clon
	4	50.2	3.8	6327	22	AAS45444	Chemically pretrea
c	5	50	3.8	10133	24	ABL32458	Human immune syste
c	6	49.4	3.8	73334	24	ABL34124	Human immune syste
c	7	49.2	3.8	11691	24	ABL34241	Human immune syste
	8	48.4	3.7	1337	20	AAZ17263	Human gene express
c	9	48	3.7	399	22	AAH93294	Plasmodium falcipa
c	10	47.8	3.7	326	22	AAS56505	Human cDNA for an
c	11	47.8	3.7	73334	24	ABL34125	Human immune syste
c	12	47.4	3.6	6627	24	ABL33574	Human immune syste
c	13	47.2	3.6	6033	21	AAA70152	Plasmodium falcipa
	14	46.8	3.6	281	22	AAL16482	Human breast cance
c	15	46.8	3.6	6107	24	AAS61315	Human gene regulat
	16	46.6	3.6	20420	22	AAK73165	Human immune/haema
c	17	46.4	3.5	463	22	AAS56574	Human cDNA for an
c	18	46	3.5	6988	24	ABL34440	Human immune syste
c	19	45.8	3.5	2140	21	AAZ49411	Pea DRR206 protein
c	20	45.8	3.5	8991	24	AAS61295	Human gene regulat
	21	45.6	3.5	3730	22	ABA19745	Human nervous syst
c	22	45.6	3.5	6699	24	AAS63321	Chemically pretrea
c	23	45.6	3.5	7049	24	ABL32157	Human immune syste
	24	45.2	3.5	13326	24	ABL33712	Human immune syste
c	25	45.2	3.5	17918	24	AAS61418	Human gene regulat
c	26	45	3.4	687	24	ABL34283	Human immune syste
	27	45	3.4	883	22	AAL26436	Human breast cance
	28	45	3.4	883	22	AAL26587	Human breast cance
	29	45	3.4	883	22	AAL26703	Human breast cance
	30	45	3.4	883	22	AAL26709	Human breast cance
	31	45	3.4	883	22	AAL26724	Human breast cance
	32	45	3.4	883	22	AAL26756	Human breast cance
	33	45	3.4	883	22	AAL26774	Human breast cance
	34	45	3.4	883	22	AAL26785	Human breast cance
c	35	45	3.4	6478	22	AAS45417	Chemically pretrea
c	36	45	3.4	6626	22	AAS46810	Tumour suppressor
	37	45	3.4	10329	24	ABL34123	Human immune syste
c	38	45	3.4	16217	24	ABL32625	Human immune syste
	39	44.8	3.4	259	22	AAL25325	Human breast cance
c	40	44.8	3.4	11473	24	ABL33355	Human immune syste
c	41	44.8	3.4	17419	22	AAS45393	Chemically pretrea
c	42	44.8	3.4	17419	24	ABL33295	Human immune syste
c	43	44.6	3.4	2286	20	AAX24385	Arabidopsis farnes
c	44	44.6	3.4	5575	24	AAS61207	Human gene regulat
c	45	44.6	3.4	5975	24	AAS61096	Human gene regulat

ALIGNMENTS

RESULT 1
AAZ35184
ID AAZ35184 standard; DNA; 12313 BP.
XX
AC AAZ35184;
XX
DT 13-MAR-2000 (first entry)
XX
DE Corn delta-12 desaturase fad2-2 gene 5' untranslated region.
XX
KW Corn; maize; transgenic plant; lipid; food; feedstuff;
KW vegetable oil; seed oil; oleic acid; fatty acid desaturase;
KW delta-12 desaturase; fad2-2; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT intron 5651..12301
FT /*tag= a
FT TATA_signal 5439..5444
FT /*tag= b
XX
PN WO9964579-A2.
XX
PD 16-DEC-1999.
XX
PF 09-JUN-1999; 99WO-US12884.
XX
PR 11-JUN-1998; 98US-0088987.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX

PI Shen JB;
XX
DR WPI; 2000-097535/08.
XX
PT New maize oleosin promoter, used for producing transgenic plants with
PT altered fatty acid composition of the oil, used e.g. in animal feeds -
XX
PS Example 2; Page 80-84; 108pp; English.
XX
CC This is the nucleotide sequence of the 5' untranslated region of
CC the novel corn fad2-2 gene that codes for delta-12 desaturase (see
CC AAY32344). The invention generally relates to the preparation and use
CC of nucleic acid fragments comprising all, or substantially all, of a
CC corn oleosin promoter (see AAZ35165-77), a stearoyl-ACP desaturase
CC (see AAZ35179-80) and a delta-12 desaturase, which can be used
CC individually or in combination to modify the lipid profile of corn.
CC Suppression of delta-12 desaturase expression can increase the
CC oleic acid content of the seed oil. Also claimed are seeds of
CC such plants, oil obtained from the grain of such plants, animal
CC feed, use of the oil in food, feed, and cooking oil or industrial
CC applications. The promoter region of fad2-2 may be used to
CC express a gene of interest in transgenic corn plants.
XX
SQ Sequence 12313 BP; 2845 A; 2969 C; 3170 G; 3329 T; 0 other;

Query Match 18.0%; Score 235; DB 21; Length 12313;
Best Local Similarity 82.5%; Pred. No. 2.5e-50;
Matches 293; Conservative 0; Mismatches 60; Indels 2; Gaps 2;

```
Qy      5 TCGCTGCTTTGTCTACATCATGTTCTTCATCATCCTCCCCAGGCGACGCGTGCTGCTGTT 64
          ||||    | ||| ||| | || | ||| ||  |  ||| ||  ||||||||||||||||
Db   3444 TCGCGTGATCGTCCACAGCTTGCTGTTCGTCGCCTACCCAAGTTGACGCGTGCTGCTGTT 3503

Qy     65 CTTATTCA-GACTACCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGT 123
          ||| |||  | | |||||||||| |||||||  ||||||| ||| ||||| ||||||  ||
Db   3504 CTTCTTCCCGGCGACCGTTCGAGGGACTGCACTGCGTACACCTTCCTGCACCGACTTCGT 3563

Qy    124 ACGGCTACATCGAACATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAA 183
          ||| ||||||||||||  | ||||||||||||||||||||||| ||| ||||  |||||| |
Db   3564 ACGACTACATCGAACAAACACACGAGATGTCTCGTGTGAATGGAGCCACTGGTGCCTTGA 3623

Qy    184 GCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTG 243
          ||||||||  |||||  |||||||  |||||||||  ||||||||||  ||| |||| ||
Db   3624 GCATCGGTCCCTCCGCTGGGTACACTCTGTTCTTCGTATTTGTGCAT-GTTTCATTGCTG 3682

Qy    244 TTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACA 303
          |||||||  ||||  ||||||||||||||  |||||||||||||  ||||||||||||||
Db   3683 TTTACTGCTTATGCGAGTAGTTATACACACATGCACATACATGTCATCACATATATCACA 3742

Qy    304 ATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGA 358
            |  ||||||  |||||||||||||||||||||| ||||  ||||||||||   |||
Db   3743 CTGATTTTCTGGATTAAATTAAAACTAAAAATGCCTAACTTTCTAACACGTCCGA 3797
```

RESULT 2
AAS96571
ID AAS96571 standard; DNA; 2126 BP.
XX
AC AAS96571;
XX
DT 26-FEB-2002 (first entry)
XX
DE Corn promoter sequence #14.
XX
KW Corn; male reproductive tissue; plant regulatory sequence; Zea mays;
KW promoter; transcription regulation; operably linked gene; monocot;
KW dicot; wheat anther; plant fertility; insect tolerance;
KW pathogen tolerance; herbicide tolerance; ds.
XX
OS Zea mays.
XX
PN WO200183790-A2.
XX
PD 08-NOV-2001.
XX
PF 30-APR-2001; 2001WO-US13739.
XX
PR 01-MAY-2000; 2000US-201255P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Conner TW, Dubois P, Malven M, Masucci JD;
XX
DR WPI; 2002-055481/07.
XX
PT Novel promoters isolated from corn for controlling gene expression in
PT male reproductive tissues, such as anthers, tassels, and to regulate
PT transcription of target genes including genes for insect or pathogen
PT tolerance -
XX
PS Claim 1; Page 115-116; 121pp; English.
XX
CC The present invention relates to the isolation of plant regulatory
CC sequences from the male reproductive tissues of corn (Zea mays). The
CC promoter sequences, fragments, regions or cis elements of the sequences,
CC are capable of regulating transcription of an operably linked DNA
CC sequence. The promoter sequences confer enhanced expression of operably
CC linked genes in monocot or dicot male reproductive tissues, such as
CC anthers, especially wheat anthers and is useful for regulating
CC transcription of a DNA sequence, by operably linking the DNA sequence
CC to the promoter. The promoter sequences are useful in plants to regulate
CC transcription of target genes including genes for control of fertility,
CC insect or pathogen tolerance and herbicide tolerance. They are also
CC useful as probes or primers in nucleic acid hybridisation experiments.
CC The promoter sequences can be used in hybridisation assays of other
CC plant tissues to identify closely related or homologous genes and
CC associated regulatory sequences. AAS96558-AAS96577 represent the
CC corn promoter sequences of the present invention.
XX
SQ Sequence 2126 BP; 607 A; 419 C; 433 G; 667 T; 0 other;

Query Match 17.8%; Score 232.8; DB 24; Length 2126;
Best Local Similarity 83.0%; Pred. No. 4.3e-50;
Matches 289; Conservative 0; Mismatches 57; Indels 2; Gaps 2;

```
Qy      5 TCGCTGCTTTGTCTACATCATGTTCTTCATCATCCTCCCCAGGCGACGCGTGCTGCTGTT 64
          ||||    | ||| ||| | || | ||| ||  |  |||| ||  ||||||||||||||||
Db     82 TCGCGTGATCGTCCACAGCTTGCTGTTCGTCGCCTTCCCAAGTTGACGCGTGCTGCTGTT 141

Qy     65 CTTATTCA-GACTACCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGT 123
          ||| |||  | | |||||||||| |||||||  |||||||||| ||||| ||||||  ||
Db    142 CTTCTTCCCGGCGACCGTTCGAGGGACTGCACTGCGTACATCTTCCTGCACCGACTTCGT 201

Qy    124 ACGGCTACATCGAACATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAA 183
          |||||||||||||||  | ||||||||||||||||||||||| ||| ||||  |||||| |
Db    202 ACGGCTACATCGAACAAACACACGAGATGTCTCGTGTGAATGGAGCCACTGGTGCCTTGA 261

Qy    184 GCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTG 243
          ||||||||  |||||  |||||||  |||||||||  ||||||||||  ||| |||| ||
Db    262 GCATCGGTCCCTCCGCTGGGTACACTCTGTTCTTCGTATTTGTGCAT-GTTTCATTGCTG 320

Qy    244 TTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACA 303
          |||||||  ||||  ||||||||||||||  |||||||||||||  ||||||||||||||
Db    321 TTTACTGCTTATGCGAGTAGTTATACACATATGCACATACATGTCATCACATATATCGCA 380

Qy    304 ATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACA 351
            |  || |||  |||||||||||||||||||| |||| |||||||||
Db    381 CTGATTATCTGGATTAAATTAAAACTAAAAATGCCTAACTTTCTAACA 428
```

RESULT 3
AAF81468/c
ID AAF81468 standard; DNA; 1888 BP.
XX
AC AAF81468;
XX

```
DT   08-JUN-2001  (first entry)
XX
DE   Corn promoter clone #700342364.
XX
KW   Corn; promoter; transgenic plant; herbicide resistance; ds.
XX
OS   Zea mays.
XX
PN   WO200119976-A2.
XX
PD   22-MAR-2001.
XX
PF   13-SEP-2000; 2000WO-US25078.
XX
PR   16-SEP-1999;   99US-0154182.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Anderson HM,  Chay CA,  Chen G,  Conner TW;
XX
DR   WPI; 2001-244796/25.
XX
PT   Novel promoter nucleic acid sequences useful for regulating
PT   heterologous gene expression in plants, comprising regulatory sequences
PT   located upstream to plant DNA structural coding sequences -
XX
PS   Claim 1; Page 95; 101pp; English.
XX
CC   The present invention relates to novel corn promoter sequences (see
CC   AAF81456-AAF81478). The promoter sequences are useful for conferring
CC   expression of a second polynucleotide molecule in a transgenic plant
CC   tissue. In addition, the promoter sequences are useful for providing
CC   plants with herbicide resistance. The promoter sequences are suitable for
CC   selectively modulating expression of any operatively linked gene and
CC   provide additional regulatory element diversity in a plant expression
CC   vector in gene stacking approaches. The present sequence is one such corn
CC   promoter sequence isolated in the present invention.
XX
SQ   Sequence 1888 BP; 569 A; 443 C; 423 G; 453 T; 0 other;

  Query Match             9.5%;  Score 124.6;  DB 22;  Length 1888;
  Best Local Similarity   63.9%;  Pred. No. 4e-22;
  Matches  227;  Conservative    0;  Mismatches  114;  Indels   14;  Gaps    2;

Qy      78 CCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGTACGGCTACATCGAA 137
           | ||| ||| |||||||  ||||| | ||  |||||||||| || ||||   || ||||||
Db     784 CTGTTTGAGGGACTGCACTGCGTAGACCTGCCTGCATCGAGTTCGTACAATTATATCGAA 725

Qy     138 CATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCC 197
           |||| |||||||||||||| |||||||| | | | || || | || |||||||||  || |
Db     724 CATACACACGAGATGTCTTGTGTGAATTGGGCCGCTGATACTTTGAGCATCGGTCCCTTC 665

Qy     198 GTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATAC 257
             |||| || | ||||| | | || || |      | |  || |||| ||||| ||||||
Db     664 ACAGGGAACTTCCTGTTTTCCATACTTTT--TGCGTGTCTTTTTTGTGTACTGCTTATAC 607

Qy     258 GAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAAT 317
           || ||||  |                 |||||| || ||||||| || |||||  ||||||  |
Db     606 GAATAGTGGT-----------ATACATGTCGTCACATAAATTACAATGGTTTTCTGGAC 559

Qy     318 TAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCA 377
           ||||||||||||  |||| || | |||||||||      ||| |   |    ||    |
Db     558 TAAATTAAAACTAGTAATGCCTCATTTTCTAACATATTTCACAGTAATAAAAGTTAGGTA 499

Qy     378 ACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACAT 432
           || ||    || |  |     | | ||  |     | | || |||  |  | ||
Db     498 TCAGCTAATTCTTTGGTGATAGGGTATAAATGACGATGCTCTATATTTATCAGAT 444


RESULT 4
AAS45444
ID   AAS45444 standard; DNA; 6327 BP.
XX
AC   AAS45444;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Chemically pretreated genomic DNA associated with cell cycle #75.
XX
KW   Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW   human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW   graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW   arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW   immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW   PCR primer.
XX
OS   Homo sapiens.
XX
PN   WO200168911-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02945.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602751/68.
XX
PT   Designing primers and probes for analysing diseases associated with
PT   cytosine methylation state e.g. arthritis, cancer, aging,
PT   arteriosclerosis comprising fragments of chemically modified genes
PT   associated with cell cycle -
XX
PS   Claim 1; SEQ ID No 149; 28pp; English.
XX
CC   Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
CC   molecules associated with the cell cycle and specific PCR primers of the
CC   invention. The sequences are useful for detecting the methylation state
CC   of all CpG dinucleotides in a sequence and therefore for analysing
CC   associated diseases. By analysing cytosine methylations in the pretreated
CC   DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC   of existing diseases or the predisposition to specific diseases can be
CC   ascertained. The parameters may be compared to another set of genetic
CC   and/or epigenetic parameters, the differences serving as basis for
CC   diagnosis and/or prognosis events which are disadvantageous to patients.
CC   The sequences of the invention are useful for the diagnosis and therapy
CC   of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC   aging, glomerular disease, Lewy body disease, arthritis,
CC   arteriosclerosis, solid tumours and cancers.
XX
SQ   Sequence 6327 BP; 1833 A; 93 C; 1407 G; 2994 T; 0 other;

  Query Match             3.8%;  Score 50.2;  DB 22;  Length 6327;
  Best Local Similarity   60.7%;  Pred. No. 0.012;
  Matches   82;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     211 TGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACA 270
           | || ||  || ||  | |  ||| ||  ||||  |||||  ||  |||| |
Db    2657 TATTGTTGATAGTTTGGTTTGAATTTTTTTAGGTTTTTGTTTTATATATATATATATATA 2716

Qy     271 TACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTA 330
           || ||  | ||| ||||    | ||||||    |||||||| |||| | ||||||||
Db    2717 TATATATATATATATATATATATATATATTTATATATTTTTGTAAAGTGAATTAAAAGAG 2776

Qy     331 AAAATGACTAAATTT 345
           | || | |||| ||
Db    2777 TATATAAATAAAGTT 2791
```

```
RESULT 5
ABL32458/c
ID   ABL32458 standard; DNA; 10133 BP.
XX
AC   ABL32458;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 431.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 431; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 10133 BP; 3075 A; 70 C; 2051 G; 4937 T; 0 other;

  Query Match             3.8%;  Score 50;  DB 24;  Length 10133;
  Best Local Similarity   51.0%;  Pred. No. 0.017;
  Matches  149;  Conservative    0;  Mismatches  135;  Indels    8;  Gaps    1;

Qy     219 TTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCA 278
           ||| |     ||| | | |      |  ||   | |||  | ||  |||  ||| || ||
Db    1416 TTAATCAAAAATACTCTAAACTAACTCAACAACTAATAAAACTACATATCTATATATACA 1357

Qy     279 CATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGAC 338
           ||||| ||  ||||||||   |  |||  | | |||||  | ||| | |  |      |
Db    1356 TATACAAATTCTCACATATTCAAATATACATATATAAATAATATTTACAAAATCCTATAT 1297

Qy     339 TAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTC----- 393
           ||||| | | | |  ||  | |    ||  |    | |    | | ||  |||||
Db    1296 TAAATATTTTAAAAAAAAAAAAAAAAAAATTCAAATACTCACATTCTAATTATTCAATAT 1237

Qy     394 ---TACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACA 450
              || ||| || || ||   |  ||  |  |||||  | |     ||||||  |||| |
Db    1236 TCTTAAATTATTATACTTAACAAATCTATAAATAAAACATAATACATACAATAAAAAAAA 1177

Qy     451 GTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATA 502
           | | ||  ||   ||||||    ||   |  |    || ||| |  | | |
Db    1176 CTTCATTTAACATTTATATATCAAATTCCTACCATATACCTAATAATAAAAA 1125


RESULT 6
ABL34124/c
ID   ABL34124 standard; DNA; 73334 BP.
XX
AC   ABL34124;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2097.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2097; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 73334 BP; 18968 A; 858 C; 15329 G; 38179 T; 0 other;

  Query Match             3.8%;  Score 49.4;  DB 24;  Length 73334;
  Best Local Similarity   53.3%;  Pred. No. 0.058;
  Matches  104;  Conservative    0;  Mismatches   91;  Indels    0;  Gaps    0;

Qy     252 TTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTT 311
           |||||    ||  |||| ||| ||  | ||| ||||    | ||||||  | |||| | |
Db   59235 TTATATATATATATATATATATATATATATATATATATATATATATATATCTATATATAT 59176

Qy     312 CTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTT 371
           |||||  || ||||   | |||| | ||||| | || ||||||  ||
Db   59175 ATAAATATAAATAAATAAATAAATAAATAAATATTTACCACCAAAAACTAAAACTAAAAA 59116

Qy     372 TCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACA 431
             || ||    |  ||| | |        |   |  | | |||    | || || ||
```

```
Db  59115 CAATTAAAAAAACAAAATATACAAACACACATTCCATTAACTATCAAAAAAAAATCATCA 59056

Qy    432 TAGTCTACAATGGAA 446
          |  ||||  ||  ||
Db  59055 TCATCTATCATATAA 59041
```

RESULT 7
ABL34241/c
ID ABL34241 standard; DNA; 11691 BP.
XX
AC ABL34241;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2214.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 2214; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 11691 BP; 3695 A; 49 C; 1861 G; 6086 T; 0 other;

  Query Match   3.8%; Score 49.2; DB 24; Length 11691;
  Best Local Similarity   51.1%; Pred. No. 0.029;
  Matches 140; Conservative   0; Mismatches 133; Indels   1; Gaps   1;

```
Qy    160 TGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCT 219
          | || | |    ||| |  |||||  ||   | | |   ||   || ||    ||   |
Db   6938 TTAAAATATAACCTATTTACTTAAATATTAATAAATATATACACTAAATATATTTACCCA 6879

Qy    220 TATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 279
          || || |  ||| |     |    ||||   || |  | | ||  |||||| | |   |
Db   6878 TAATTCTAAATACTACATATACATTTTAAATATAACTC-ATTAAATATACACAAAAAAAT 6820

Qy    280 ATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACT 339
          ||| ||||   ||||||  |  |  |||    | || ||  ||   | ||  | |||   |
Db   6819 ATATATATATACACATAAAATAACATAAAAATTTACATATAAAATACACATATAATATTT 6760

Qy    340 AAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATT 399
          |    ||||  |  ||  ||||  |||   ||| |  |  || | ||  ||   |  ||
Db   6759 ACTAATCTATAAATAAATACATATTAACACTTTATTTATTAAATATATATAAATTCTATA 6700

Qy    400 GTTCTATTTCGAATTTCACTCTATAAACAACATA 433
          |||  |     | || |||  || ||| || | |
Db   6699 TTTCAAAAATAATTTACACAATAAAAAAAATACA 6666
```

RESULT 8
AAZ17263
ID AAZ17263 standard; cDNA; 1337 BP.
XX
AC AAZ17263;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:4735.
XX
KW Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
PN WO9938972-A2.
XX
PD 05-AUG-1999.
XX
PF 28-JAN-1999; 99WO-US01619.
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
DR WPI; 1999-494092/41.
XX
PT Novel human genes and their expression products which are
PT differentially expressed in different cell types
XX
PS Claim 1; Page 2250-2251; 2479pp; English.
XX
CC The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as

```
CC   cancer). The polynucleotides of the invention are especially used in the
CC   diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC   and lung cancer. The polynucleotides can also be used to screen for
CC   peptide analogues and antagonists.
XX
SQ   Sequence 1337 BP; 42 A; 577 C; 27 G; 22 T; 669 other;

  Query Match             3.7%;  Score 48.4;  DB 20;  Length 1337;
  Best Local Similarity   34.5%;  Pred. No. 0.018;
  Matches   97;  Conservative    0;  Mismatches  184;  Indels    0;  Gaps    0;

Qy    785 CTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACGTCCACGGCACCA 844
Db    381 CCNNNNCCNCNNNNNNCNCNNNNNNNNNCCCCCCNCNCCNNCNCCCCNNNCNCCCNNNNN 440

Qy    845 TCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAACCACCCCCAACC 904
Db    441 NCCNCCCCCNCNNNNNCNNNNCCCCCCCCCNNNNNNNNCNNCCCCNNCCNCCCCCCNCCC 500

Qy    905 CACTCACGAAACCCCGTCCCGGCCGTGCCCGTGTCGGTCCGCGCTCGGCAACGAGGCGGC 964
Db    501 CNCNCCCCCCCCCCCCCCCCCNCCCNNCCCCNNCCCNCCCCNCNCNCNCCCCCNNNCCCNN 560

Qy    965 CCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTGTCGGCCCTTTGTTTATTCATCCCG 1024
Db    561 CCCCNCCNNNNCCCCCCNCCNCCCCCCCCNNCCNCNCNCCCCCCNCCCCCCCCCNCNNNCCC 620

Qy   1025 AAATCTCATCTGCCCCCACGGCCGACTGCGCTGCGCCGCCC 1065
Db    621 CCCNCCNCNCCCCCCCCCCCCCCNNNNCCCCCCCCCCCCCC 661


RESULT 9
AAH93294/c
ID   AAH93294 standard; DNA; 399 BP.
XX
AC   AAH93294;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Plasmodium falciparum MAL3P8 polynucleotide SEQ ID NO 16.
XX
KW   Human; antisense-therapy; gene-therapy; diagnostic; forensic;
KW   gene mapping; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200152616-A2.
XX
PD   26-JUL-2001.
XX
PF   22-DEC-2000; 2000WO-US35190.
XX
PR   23-DEC-1999;   99US-0471275.
PR   21-JAN-2000; 2000US-0488725.
PR   25-APR-2000; 2000US-0552317.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-451890/48.
XX
PT   Isolated polypeptide for treatment of diseases, diagnostics, raising
PT   antibodies and research use -
XX
PS   Example 4; Page 106; 135pp; English.
XX
CC   The invention relates to an isolated human polynucleotide (AAH75398)
CC   encoding a novel polypeptide (AAG64527) useful in antisense-therapy and
CC   gene-therapy, in diagnostics, forensics, gene mapping and identification
CC   of mutations responsible for genetic disorders and other traits.
CC   Polynucleotide sequences with potential homology were also identified
CC   (AAH93283-AAH93356).
XX
SQ   Sequence 399 BP; 225 A; 24 C; 15 G; 135 T; 0 other;

  Query Match             3.7%;  Score 48;  DB 22;  Length 399;
  Best Local Similarity   47.1%;  Pred. No. 0.013;
  Matches  147;  Conservative    0;  Mismatches  165;  Indels    0;  Gaps    0;

Qy    208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
Db    320 TTTTTTTTTTTTATATATTTGTCAATTTTTATGTTTGCATAAACGATTAATATATATCTTT 261

Qy    268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
Db    260 TTTTACATTATAATATATATAAATATATATATATATATATATATATATATATATATAT 201

Qy    328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
Db    200 ATATATATTTATTTATTTATTTTACGATAGTTTTAATTTTTTTTTTTTATCTGCGAATTA 141

Qy    388 CTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAA 447
Db    140 AACATTTTAATTATGTTCTTTATATTAATTATATATAATATATTATATAAAAATAAAA 81

Qy    448 ACAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAGTCGT 507
Db     80 TTAGAATGTTTTGATATTTTTAAGGTTTTTTTTTTTTTTTTTTGATACACAATTCATTA 21

Qy    508 TTGAAGATTGAA 519
Db     20 TTTAAATTATAA 9


RESULT 10
AAS56505/c
ID   AAS56505 standard; cDNA; 326 BP.
XX
AC   AAS56505;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human cDNA for an ovarian cancer protein #129.
XX
KW   Human; ss; ovarian cancer protein; cancer; tumour; ovarian cancer;
KW   endometrial cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200170976-A2.
XX
PD   27-SEP-2001.
XX
PF   20-MAR-2001; 2001WO-US09062.
XX
PR   21-MAR-2000; 2000US-190710P.
PR   22-JUN-2000; 2000US-213748P.
PR   19-DEC-2000; 2000US-257276P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Pyle RA,  Stolk JA;
XX
DR   WPI; 2001-607531/69.
XX
PT   Nucleic acids encoding 222 polypeptides associated with ovarian and
PT   endometrial cancers, useful for diagnosing, preventing and treating
PT   cancers -
XX
PS   Claim 1; Page 157; 187pp; English.
XX
CC   The invention relates to human polynucleotides encoding proteins
CC   associated with ovarian and endometrial cancers. The polynucleotides and
CC   the proteins they encode may be used in the prevention, diagnosis and
```

CC treatment of diseases associated with the inappropriate expression of
CC ovarian and endometrial cancer polypeptides (OECPs). For example, the
CC polynucleotide (or an expression vector comprising the polynucleotide)
CC and the OECP may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of OECPs by expressing inactive proteins or to
CC supplement the patients own production of them. Additionally, the
CC polynucleotide may be used to produce the OECPs, by inserting the nucleic
CC acids into a host cell and culturing the cell to express the protein. The
CC polynucleotide and its complementary sequences may also be used as DNA
CC probes in diagnostic assays to detect and quantitate the presence of
CC similar nucleic acids in samples, and therefore which patients may be in
CC need of restorative therapy. The OECPs may also be used as antigens in
CC the production of anti-OECP antibodies and in assays to identify
CC modulators of it's expression and activity. The anti-OECP antibodies and
CC antagonists may also be used to down regulate expression and activity.
CC The anti-OECP antibodies may also be used as diagnostic agents for
CC detecting the presence of OECP in samples (e.g. by enzyme linked
CC immunosorbant assay (ELISA)) and hence diagnose patients with
CC cancers. The present sequence is a ovarian and endometrial cancer linked
CC cDNA of the invention.
XX
SQ Sequence 326 BP; 130 A; 48 C; 34 G; 114 T; 0 other;

Query Match 3.7%; Score 47.8; DB 22; Length 326;
Best Local Similarity 55.0%; Pred. No. 0.014;
Matches 94; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
Qy    211 TGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACA 270
          | || || || ||| |   | |||      |||   ||   ||  |  | |||  |||||
Db    326 TTTTTTTTTTTTTTTTTTTTTTTTCAGGGATTGAACACATTTTTAATTATTAGAAATACA 267

Qy    271 TACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTA 330
          |||||||| |||  || | |   || |||    | |  |   | |     || ||||  |
Db    266 TACATGCATATAATTAACCTTTAATTTATTTTTAAAAATAAATTATGGTCATGAAAAACA 207

Qy    331 AAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAA 381
          | | ||  |||| ||  ||   ||  |||||   | | | ||| | || ||
Db    206 ATAGTGCTTAAAATTGAAATTTCATAGACATGTAATTTTATTCCCTAAAAA 156
```

RESULT 11
ABL34125/c
ID ABL34125 standard; DNA; 73334 BP.
XX
AC ABL34125;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 2098.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 2098; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 73334 BP; 24605 A; 858 C; 14432 G; 33439 T; 0 other;

Query Match 3.7%; Score 47.8; DB 24; Length 73334;
Best Local Similarity 51.1%; Pred. No. 0.15;
Matches 112; Conservative 0; Mismatches 107; Indels 0; Gaps 0;

```
Qy    232 TTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATC 291
          | |||||   | || | | ||||    | ||   | | |||||| || | | |   | |
Db  66278 TATTTATAAATCTTAAATAATTAACTTAATATAAAAAAATACATACAAAAAAAACACCTT 66219

Qy    292 ACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACA 351
          | | || | ||||||||||  ||||   |   |||   |||||  |  | | ||| || |
Db  66218 AAAAATTTTACAATATTTTAATAAAAATACAAAAATAAAAAAAAAAAAACAATTCAAATA 66159

Qy    352 CCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGA 411
          |||  | |   ||||   || | |   | |       | | || |       |     |
Db  66158 ACAAACAAAACATAATAACTTATACCTATAATCCCAACACTTTAAAAAACGAAAACGAAA 66099

Qy    412 ATTTCACTCTATAAACAACATAGTCTACAATGGAAAACA 450
          |  |||||  ||    || ||      | |  | |||||
Db  66098 AAATCACTTAATCCCAAAAATTCGAATCCAACGTAAACA 66060
```

RESULT 12
ABL33574/c
ID ABL33574 standard; DNA; 6627 BP.
XX
AC ABL33574;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1547.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.

```
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 1547; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 6627 BP; 2052 A; 117 C; 1386 G; 3072 T; 0 other;

  Query Match             3.6%;  Score 47.4;  DB 24;  Length 6627;
  Best Local Similarity   49.2%;  Pred. No. 0.066;
  Matches  155;  Conservative    0;  Mismatches  156;  Indels    4;  Gaps    1;

Qy     192 TACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGA 251
           |||||  ||   ||| | ||  |  ||   |   | |||  |   |||     | | | |
Db    6510 TACTCTTTAAAATACCTACTAATTCTCAATTCCTTTCATCCTAACATTAACCCTAAATAA 6451

Qy     252 TTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTT 311
           ||| |    || | ||  | | ||  | | ||  |  | || | | || | |  | |||
Db    6450 TTACAATTATATTAATTTACAAATAAAAAAACTAAAAACCAAAAAAATTAAATAACTTTA 6391

Qy     312 CTAAATT----AAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAAT 367
           |||| ||     || |||||  ||||  |||||| ||  ||| | ||  ||      ||
Db    6390 CTAACTTTACCCAACTAAAAAATAAAAAAACTAAAATTTAAACCCAAAAAACTAAACCAT 6331

Qy     368 GTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAAC 427
           |  |||  || | ||  ||     |  ||||  |  |  |  | | ||| |  ||
Db    6330 ATAATCTAAAAAACCTAAACTCCATAATCATTACTAAACATAAACTATCAATTAATCCTT 6271

Qy     428 AACATAGTCTACAATGGAAAACAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACA 487
           |  | |      ||  | ||||| | || | | |   |  | |  | |         |||
Db    6270 ATAACAACTCCAAAACGCAAACAATACTATCTCCATTTCAAAAACCAACAAACTAAAACA 6211

Qy     488 ACATAAGACAATATA 502
            ||   || ||  ||
Db    6210 CCAATCGAAAAACTA 6196


RESULT 13
AAA70152/c
ID   AAA70152 standard; DNA; 6033 BP.
XX
AC   AAA70152;
XX
DT   07-NOV-2000  (first entry)
XX
DE   Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:285.
XX
KW   Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KW   antimalarial; malaria; protozoacide; infection; insecticide; ds.
XX
OS   Plasmodium falciparum.
XX
PN   WO200025728-A2.
XX
PD   11-MAY-2000.
XX
PF   05-NOV-1999;   99WO-US26796.
XX
PR   05-NOV-1998;   98US-0107131.
XX
PA   (HOFF/) HOFFMAN S.
PA   (CARU/) CARUCCI D.
PA   (GARD/) GARDNER M.
PA   (VENT/) VENTER J C.
XX
PI   Hoffman S,  Carucci D,  Gardner M,  Venter JC;
XX
DR   WPI; 2000-365347/31.
XX
PT   Proteins encoded by chromosome 2 of the human malarial parasite,
PT   Plasmodium falciparum, useful as antimalarial vaccines and in the
PT   diagnosis of P.falciparum infection -
XX
PS   Disclosure; Page 493-495; 577pp; English.
XX
CC   The present invention describes proteins and their fragments (I) encoded
CC   by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC   Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
CC   vaccines against P. falciparum infection comprising (I) or (II).
CC   (I) and (II) are useful for the development of vaccines against
CC   P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC   antibody raised to immunogens comprising the sequences of (I), are
CC   useful in the detection of infection with P. falciparum. Furthermore,
CC   (I) (especially when they are rifins or secreted or membrane proteins)
CC   can aid the identification of drugs to treat or prevent P. falciparum
CC   infection, or they can be used to identify drug resistance in
CC   P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
CC   subsequent identification of proteins encoded by it will help to expand
CC   our understanding of parasite biology, a process hampered by the
CC   complexity of the parasitic lifecycle, and provide new targets for
CC   vaccine and drug development. Parasite resistance to drugs and mosquito
CC   resistance to insecticides have led to a resurgence of malaria in many
CC   parts of the world, and there is a pressing need for vaccines and new
CC   drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC   and protein sequences given in the present invention, but which are not
CC   specifically mentioned within the specification.
XX
SQ   Sequence 6033 BP; 3019 A; 437 C; 707 G; 1870 T; 0 other;

  Query Match             3.6%;  Score 47.2;  DB 21;  Length 6033;
  Best Local Similarity   49.6%;  Pred. No. 0.071;
  Matches  121;  Conservative    0;  Mismatches  123;  Indels    0;  Gaps    0;

Qy     205 ACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGT 264
           ||| ||| ||||   ||       || | | || | ||  |||   ||  ||    || |
Db    2523 ACAATCTTTTCTATGTACAACCATATTTATATAATATTTATTATATATATTAGTTATACT 2464

Qy     265 TATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTA 324
             || ||  ||   ||    | ||| ||  | |||   || ||||  ||| |||||  |
Db    2463 GTTAAATGTATTATCAATATTTTCAACATTTTTATTTAAAAATTTAAATAATTTAAAAAA 2404

Qy     325 AAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTT 384
           |||| | | | ||| | ||   |  |   |||| ||  | || | ||    |
Db    2403 TTTTGAAAATTCATTCAATATATATTTCATATATAATTGAAAGATATTATACACATTCCA 2344

Qy     385 TACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGG 444
           | | | |  | | || |  | |||     ||   | | | |      ||    | ||| |
Db    2343 TTCTTCTAATCCTTTTTCATTTTTTTGTCTTTTTTTTTTTTATATTATTATAAAAAAATTG 2284

Qy     445 AAAA 448
            |||
Db    2283 GAAA 2280


RESULT 14
AAL16482
ID   AAL16482 standard; cDNA; 281 BP.
XX
AC   AAL16482;
XX
DT   07-DEC-2001  (first entry)
XX
```

DE Human breast cancer expressed polynucleotide 8939.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 1617; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX
SQ Sequence 281 BP; 93 A; 32 C; 39 G; 117 T; 0 other;

Query Match 3.6%; Score 46.8; DB 22; Length 281;
Best Local Similarity 54.7%; Pred. No. 0.023;
Matches 93; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 212 GTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACAT 271
Db 9 GTTTTTTTTTTTTTTTTTTTTTTCAGGGTTTGAACACATTTTTAATTATTAGAAATACAT 68

Qy 272 ACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAA 331
Db 69 ACATGCTTATAATTAACCTTAAATTTATTTTTAAAAATAAATTATGGTCATGAAAAACAA 128

Qy 332 AAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAA 381
Db 129 TATGGCTTAAAATTGAAATTTCATAGACATGTAATTTTATTCCCTAAAAA 178

RESULT 15
AAS61315/c
ID AAS61315 standard; DNA; 6107 BP.
XX
AC AAS61315;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #270.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP03968.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
PT with gene regulation, useful for analysing cytosine methylations for
PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
PT disease -
XX
PS Disclosure; SEQ ID No 276; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The
CC chemical pretreatment converts cytosine bases unmethylated at the
CC 5-position to uracil or another base with hybridisation behaviour
CC dissimilar to cytosine, to enable analysis of cytosine methylations.
CC The DNA sequences, oligomers (or sets/arrays) and method are
CC useful in the diagnosis of diseases (or predisposition to diseases)
CC associated with gene regulation and in therapy of such diseases, by
CC enabling analysis of the cytosine methylation patterns of such genes,
CC kits are provided. They are especially useful in diagnosis
CC and therapy of e.g. severe combined immunodeficiency disease, cardiac
CC disorders, haemophilia, solid tumours and cancer, Werner syndrome,
CC asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease,
CC preeclampsia, graft versus-host disease. The present sequence is a
CC sequence included in the sequence data for this specification and is
CC associated with the human gene regulation-associated genes.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6107 BP; 1822 A; 187 C; 1337 G; 2761 T; 0 other;

Query Match 3.6%; Score 46.8; DB 24; Length 6107;
Best Local Similarity 49.2%; Pred. No. 0.091;
Matches 123; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 220 TATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 279
Db 2615 TTTATATATATATATATTTATATATATATTTATTACATTTTTATTTCCACAAACTCTATC 2556

Qy 280 ATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACT 339
Db 2555 ATAAATCTAATCATTAATAACATATATTTAATTAAAAAAATATATAACTTAATAAATTAT 2496

Qy 340 AAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATT 399
Db 2495 ATAAATATTATATTAAAATAAATTTCACATTATCTTAATTTACAAAAACTAAAATATCTA 2436

Qy 400 GTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGT 459
Db 2435 AAAAACTAAAAAATATCAATCAACAAACTATTATTTTAACAATTATACATAACACATAAA 2376

```
Qy     460 ACGACTATAT 469
           |   ||| ||
Db    2375 AATTCTAAAT 2366
```

Search completed: November 7, 2002, 10:50:17
Job time : 319.436 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 07:44:03 ; Search time 38.9851 Seconds
(without alignments)
8247.617 Million cell updates/sec

Title: US-09-905-558C-16
Perfect score: 1309
Sequence: 1 cccatcgctgctttgtctac.....caagggaaagtgatcccatg 1309

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	51.2	3.9	7218	1	US-08-232-463-14	Sequence 14, Appl
c	2	43.4	3.3	2781	3	US-08-749-522-4	Sequence 4, Appli
	3	43	3.3	796	4	US-09-007-119-14	Sequence 14, Appl
c	4	43	3.3	809	4	US-09-007-119-10	Sequence 10, Appl
c	5	41.8	3.2	1890	6	5312912-3	Patent No. 5312912
c	6	41.6	3.2	1441	4	US-08-821-994-63	Sequence 63, Appl
c	7	41	3.1	4254	2	US-08-443-639-7	Sequence 7, Appli
c	8	40.6	3.1	6152	4	US-08-973-462-1	Sequence 1, Appli
	9	40.4	3.1	1000	1	US-08-599-252-96	Sequence 96, Appl
	10	40.4	3.1	1000	5	PCT-US96-06352-96	Sequence 96, Appl
	11	40.4	3.1	1000	5	PCT-US96-06583-96	Sequence 96, Appl
c	12	40.4	3.1	10342	4	US-08-972-927-5	Sequence 5, Appli
	13	40	3.1	4673	1	US-07-638-431-1	Sequence 1, Appli
	14	40	3.1	4673	5	PCT-US92-00018-1	Sequence 1, Appli
c	15	40	3.1	5829	4	US-09-004-838-109	Sequence 109, App
c	16	39.6	3.0	51952	3	US-08-947-823-1	Sequence 1, Appli
c	17	39.2	3.0	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	18	39	3.0	320	4	US-09-165-264-11	Sequence 11, Appl
c	19	38.8	3.0	923	4	US-09-004-731-66	Sequence 66, Appl
c	20	38.8	3.0	923	4	US-08-749-699-66	Sequence 66, Appl
c	21	38.6	2.9	320	4	US-09-165-264-13	Sequence 13, Appl
c	22	38.4	2.9	1368	3	US-08-874-563-5	Sequence 5, Appli
c	23	38.4	2.9	1368	3	US-08-577-483-14	Sequence 14, Appl
c	24	38.4	2.9	1736	3	US-09-182-816-22	Sequence 22, Appl
	25	38.4	2.9	1736	3	US-09-182-816-24	Sequence 24, Appl
c	26	38.4	2.9	1736	3	US-09-471-528-22	Sequence 22, Appl
	27	38.4	2.9	1736	3	US-09-471-528-24	Sequence 24, Appl
c	28	38.4	2.9	1736	4	US-09-634-530-22	Sequence 22, Appl
	29	38.4	2.9	1736	4	US-09-634-530-24	Sequence 24, Appl
c	30	38.4	2.9	4253	3	US-08-577-483-7	Sequence 7, Appli
	31	38.2	2.9	4526	1	US-07-855-412B-4	Sequence 4, Appli
	32	38.2	2.9	4526	2	US-08-308-887A-4	Sequence 4, Appli
	33	38.2	2.9	4526	3	US-08-881-094-4	Sequence 4, Appli
c	34	38	2.9	319	4	US-09-165-264-8	Sequence 8, Appli
	35	37.8	2.9	611	4	US-09-385-982-393	Sequence 393, App
c	36	37.8	2.9	899	1	US-07-820-154A-3	Sequence 3, Appli
c	37	37.8	2.9	899	2	US-08-097-554A-3	Sequence 3, Appli
c	38	37.8	2.9	899	3	US-08-480-640A-3	Sequence 3, Appli
c	39	37.8	2.9	899	3	US-08-295-802-3	Sequence 3, Appli
c	40	37.8	2.9	899	4	US-08-488-237A-3	Sequence 3, Appli
c	41	37.8	2.9	899	4	US-08-375-992A-3	Sequence 3, Appli
c	42	37.8	2.9	899	5	PCT-US93-00324-3	Sequence 3, Appli
	43	37.8	2.9	925	3	US-08-858-003-1	Sequence 1, Appli
	44	37.8	2.9	925	3	US-09-078-166-1	Sequence 1, Appli
	45	37.8	2.9	925	4	US-08-997-467-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-Fls
US-08-232-463-14

  Query Match                 3.9%;  Score 51.2;  DB 1;  Length 7218;
```

```
  Best Local Similarity   4.3%;  Pred. No. 0.00018;
  Matches   17;  Conservative  219;  Mismatches  162;  Indels    0;  Gaps    0;

Qy     791 CGGCAAATCAAACGAGGTGGTGCCCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTC 850
           | |||  || |  ||| | | |  :: : ::: :::::: : ::: :  : :: ::::::
Db    1043 CTGCAGGTCGAGGGAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1102

Qy     851 TGCAGCCGCTCACCAGCCATGCCGTGTCGCGGAACGGCACAACCACCCCCAACCCACTCA 910
           : :  :: ::: ::  :: : :: : :: :    :  : :  :: :::::  ::: :::
Db    1103 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1162

Qy     911 CGAAACCCCGTCCCGGCCGTGCCCGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGC 970
           :    :::: ::::  :: : ::: : ::  ::: : :::  :  :    :  ::: : :
Db    1163 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1222

Qy     971 TGCTGAGTCCCCTGGACACCCGACACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCT 1030
           : ::   ::::::   : :::  : :::: ::  :::::: ::: ::: ::::    :::
Db    1223 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1282

Qy    1031 CATCTGCCCCCACGGCCGACTGCGCTGCGCCGCCCGGATATATATACCCATCGTTATCGA 1090
           : ::: ::::: :  ::  :: : :: : :: :::   : : : : ::: :: :: ::
Db    1283 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1342

Qy    1091 TCGATCGATCGCGTCACTCACGGGTAGCTCATGGTCGAGCGTAGCATGCAGGAACTTATT 1150
           ::  ::  :: : :: ::: :   :  ::: :  ::   : :  : : :     ::: ::
Db    1343 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1402

Qy    1151 TGCCGTGCGCTCCCAGGTCTCCGCTCGCGTGCCTTCCA 1188
           : :: : : :::::   ::::: ::: : : : | |||
Db    1403 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCA 1440


RESULT 2
US-08-749-522-4/c
; Sequence 4, Application US/08749522
; Patent No. 6096950
;  GENERAL INFORMATION:
;    APPLICANT:  John, Maliyakal
;    TITLE OF INVENTION:  FIBER-SPECIFIC PROMOTERS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  411 East Wisconsin Avenue
;      CITY:  Milwaukee
;      STATE:  WI
;      COUNTRY:  U.S.A.
;      ZIP:  53202-4497
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/749,522
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Baker, Jean C.
;      REGISTRATION NUMBER:  35,433
;      REFERENCE/DOCKET NUMBER:  670513.90244
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (414) 277-5709
;      TELEFAX:  (414) 271-3552
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2781 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-749-522-4

  Query Match             3.3%;  Score 43.4;  DB 3;  Length 2781;

  Best Local Similarity   50.7%;  Pred. No. 0.018;
  Matches  104;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy     228 CATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATAT 287
           ||   ||| || ||| |||| |    || |||    ||||| ||   |  |  || ||||
Db    1343 CAGGATTTAATAGTTATTTAGTTTAGATTCGAAATTTTATATATTATTTTATTTATATAT 1284

Qy     288 CATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCT 347
           | | |||||| | |||  | ||  |||| |  |  |   || ||||    | || |
Db    1283 TTTAATATATATTATAATTATATTTAAAATAATTTATATTTTATAAATTTTGATATATTA 1224

Qy     348 AACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATT 407
           |  |   |  | ||   ||| ||   |         |||   |||| ||      ||   |
Db    1223 ATAATATATTATATATAAAATTTTAATTTATTTTTATTTTTATATTATATTAATATCATAT 1164

Qy     408 TCGAATTTCACTCTATAAACAACAT 432
           |  |||||   |   || |  | ||
Db    1163 TTTAATTTATTTTATTATATCATAT 1139


RESULT 3
US-09-007-119-14
; Sequence 14, Application US/09007119C
; Patent No. 6300541
; GENERAL INFORMATION:
;  APPLICANT: Lighfoot, David A.
;  APPLICANT: Gibson, Paul T.
;  APPLICANT: Merkem, Khalid
;  TITLE OF INVENTION: Soybean Sudden Death Syndrome Resistant Soybeans,
;  TITLE OF INVENTION: Soybean Cyst Nematode Resistant Soybeans and Methods of
;  TITLE OF INVENTION: Breeding and Identifying Resistant Plants
;  FILE REFERENCE: Sou Illinois 1268/2 Sequence Listing
; Patent No. 6300541
;  CURRENT APPLICATION NUMBER: US/09/007,119C
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,335
;  EARLIER FILING DATE: 1997-01-14
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 796
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (3)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (20)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (473)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (479)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (610)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (668)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (677)..(678)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
```

```
;   LOCATION: (710)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (723)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (771)..(772)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (787)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (792)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (796)
;   OTHER INFORMATION: (a or c or g or t/u)
US-09-007-119-14

  Query Match             3.3%;  Score 43;  DB 4;  Length 796;
  Best Local Similarity   54.8%;  Pred. No. 0.012;
  Matches   85;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     221 ATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACA 280
           ||||||| |  ||  || |    ||       ||||  || |   |||  ||  | ||||
Db     148 ATTTGTGAAATTTGGTACTCCCTTTGGACTCGTATATAAGAAAAAATAACTAATTTCACA 207

Qy     281 TACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTA 340
           |  ||       |  ||  |||||  ||||   |  | |||  | |||  ||||||  |
Db     208 TTAATTAAGAAAGTTAGTTCACATCATTTAATTTTACTAATCTCAAATAAAAAATAAGGT 267

Qy     341 AATTTCTAACACCAACGACATTGTAATGTTTTCTC 375
            ||||||| |    |  ||||| ||  | || ||
Db     268 TATTTCTAAAATGTCCTTCATTGGAACTTGTTATC 302


RESULT 4
US-09-007-119-10/c
; Sequence 10, Application US/09007119C
; Patent No. 6300541
; GENERAL INFORMATION:
;  APPLICANT: Lighfoot, David A.
;  APPLICANT: Gibson, Paul T.
;  APPLICANT: Merkem, Khalid
;  TITLE OF INVENTION: Soybean Sudden Death Syndrome Resistant Soybeans,
;  TITLE OF INVENTION: Soybean Cyst Nematode Resistant Soybeans and Methods of
;  TITLE OF INVENTION: Breeding and Identifying Resistant Plants
;  FILE REFERENCE: Sou Illinois 1268/2 Sequence Listing
; Patent No. 6300541
;  CURRENT APPLICATION NUMBER: US/09/007,119C
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,335
;  EARLIER FILING DATE: 1997-01-14
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 809
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(6)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (506)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (619)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (662)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (687)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (747)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (766)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (769)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (785)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (795)
;   OTHER INFORMATION: (a or c or g or t/u)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (797)
;   OTHER INFORMATION: (a or c or g or t/u)
US-09-007-119-10

  Query Match             3.3%;  Score 43;  DB 4;  Length 809;
  Best Local Similarity   54.8%;  Pred. No. 0.012;
  Matches   85;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy     221 ATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACA 280
           ||||||| |  ||  || |    ||       ||||  || |   |||  ||  | ||||
Db     457 ATTTGTGAAATTTGGTACTCCCTTTGGACTCGTATATAAGAAAAAATAACTAATTTCACA 398

Qy     281 TACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTA 340
           |  ||       |  ||  |||||  ||||   |  | |||  | |||  ||||||  |
Db     397 TTAATTAAGAAAGTTAGTTCACATCATTTAATTTTACTAATCTCAAATAAAAAATAAGGT 338

Qy     341 AATTTCTAACACCAACGACATTGTAATGTTTTCTC 375
            ||||||| |    |  ||||| ||  | || ||
Db     337 TATTTCTAAAATGTCCTTCATTGGAACTTGTTATC 303


RESULT 5
5312912-3/c
;Patent No. 5312912
;    APPLICANT: HADWIGER, LEE A.;CHIANG, CHIN C.;HOROVITZ,
;DANIEL A.
;    TITLE OF INVENTION: PROCEDURES AND REGULATORY DNA SEQUENCES
;FOR GENETICALLY ENGINEERING DISEASE RESISTANCE AND OTHER
;INDUCIBLE TRAITS IN PLANTS
;    NUMBER OF SEQUENCES: 9
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/393,301
;      FILING DATE: 13-JUN-1989
;SEQ ID NO:3:
;      LENGTH: 1890
5312912-3

  Query Match             3.2%;  Score 41.8;  DB 6;  Length 1890;
  Best Local Similarity   53.3%;  Pred. No. 0.041;
  Matches   88;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;
```

```
Qy    232 TTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACATATCATC 291
          ||| | || |   ||       | || |  |   ||| ||| ||  | ||| ||||
Db    288 TTTATTTTTTAAATTTTCACACACACAAAAAAAAATATATATATATATATATATATATAT 229

Qy    292 ACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACA 351
          | ||||||    |||| | | || ||  |  || |  || | || |  || |   || |
Db    228 ATATATATATATATATATATATATATATATATATATATATATATATTAACAACTCAATATCT 169

Qy    352 CCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTAC 396
          |   |   ||| |  | |||| |  ||| |||| |||  |   ||
Db    168 CTGTCATAATTATTTTCTTTTTTTTAACTACTTAACCCTTATGAC 124


RESULT 6
US-08-821-994-63/c
; Sequence 63, Application US/08821994A
; Patent No. 6228643
; GENERAL INFORMATION:
;  APPLICANT: Greenland, Andrew J
;  APPLICANT: Thomas, Didier RP
;  APPLICANT: Jepson, Ian
;  TITLE OF INVENTION: Promoters
;  FILE REFERENCE: PPD 50108
;  CURRENT APPLICATION NUMBER: US/08/821,994A
;  CURRENT FILING DATE: 1997-03-22
;  EARLIER APPLICATION NUMBER: PCT/GB97/00729
;  EARLIER FILING DATE: 1997-03-18
;  EARLIER APPLICATION NUMBER: GB 9606062.9
;  EARLIER FILING DATE: 1996-03-22
;  NUMBER OF SEQ ID NOS: 89
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 63
;   LENGTH: 1441
;   TYPE: DNA
;   ORGANISM: Brassica napus
US-08-821-994-63

  Query Match             3.2%;  Score 41.6;  DB 4;  Length 1441;
  Best Local Similarity   48.7%;  Pred. No. 0.04;
  Matches  113;  Conservative    0;  Mismatches  119;  Indels    0;  Gaps    0;

Qy    211 TGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACA 270
          | || || || ||| |   | ||||| || || |||  |  || |     |  || |
Db   1441 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1382

Qy    271 TACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTA 330
          |   |  | | | |||||||      |||     | ||  |  | | | ||   |||  |
Db   1381 TTTTTTGAGAAATATATCATTTTTATTATATTTCTCTTAATACATAATCAAGCCAAAGCA 1322

Qy    331 AAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTA 390
          || ||   |  | |   |  |       ||   ||  | ||| | ||       | ||||
Db   1321 AATATTCTTTGAATGAAATGATATTGACCAACATAGAGATTTATACAGATTAAATCCCTA 1262

Qy    391 TTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAAT 442
          |  |  ||   | | ||||||| | |  |  ||| | |  ||| |  |  ||
Db   1261 TGGTGAATCAATTTTTTTCGAAATACTATACATACAAACAATAATACAAGAT 1210


RESULT 7
US-08-443-639-7/c
; Sequence 7, Application US/08443639
; Patent No. 5981843
;  GENERAL INFORMATION:
;     APPLICANT:  Chappell, Joseph
;     APPLICANT:  Yin, Shaohui
;     APPLICANT:  Cornett, Catherine A.G.
;     TITLE OF INVENTION:  Transcriptional Control Sequences and
;     TITLE OF INVENTION:  Methods
;     NUMBER OF SEQUENCES:  13
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Greenlee and Winner, P.C.
;       STREET:  5370 Manhattan Circle, Suite 201
;       CITY:  Boulder
;       STATE:  CO
;       COUNTRY:  US
;       ZIP:  80303
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/443,639
;       FILING DATE:  18-MAY-1995
;       CLASSIFICATION:  800
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Ferber, Donna M.
;       REGISTRATION NUMBER:  33878
;       REFERENCE/DOCKET NUMBER:  69-94
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (303) 499-8080
;       TELEFAX:  (303) 499-8089
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4254 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     ORIGINAL SOURCE:
;       ORGANISM:  Nicotiana tabacum
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  join(1217..1327, 1455..1718, 1806..2182,
;       LOCATION:  2259
;       LOCATION:  ..2477, 2609..2747, 2903..3148, 3262..3558)
US-08-443-639-7

  Query Match             3.1%;  Score 41;  DB 2;  Length 4254;
  Best Local Similarity   52.7%;  Pred. No. 0.1;
  Matches   89;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy    265 TATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTA 324
          | |||| |||  |||| |||||   | | ||||||    |||| | | || ||  |  ||
Db    367 TGTACACACACACACACACATACATTTATATATATATATATATATATATATATATATA 308

Qy    325 AAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTT 384
          |  || | ||   || || | ||  |         ||| |  |    | | ||  |  ||
Db    307 TATATATATATATATATATATATTAATTCTATAATTTTCTTTAACTTTTCACAATTTT 248

Qy    385 TACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATA 433
          |||| | |    || |||   || |||     | || | || |||
Db    247 AGTGTATTTTTCTGATTTATATAATGATTTTATAAATTTATAAAATATA 199


RESULT 8
US-08-973-462-1/c
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT: DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 0660-0125-0 PCT
;  CURRENT APPLICATION NUMBER: US/08/973,462B
;  CURRENT FILING DATE: 1998-02-06
;  EARLIER APPLICATION NUMBER: PCT/FR96/00894
;  EARLIER FILING DATE: 1996-06-12
;  EARLIER APPLICATION NUMBER: FR 95/07007
;  EARLIER FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
```

```
;   LENGTH: 6152
;   TYPE: DNA
;   ORGANISM: P. falciparum
US-08-973-462-1

  Query Match             3.1%;  Score 40.6;  DB 4;  Length 6152;
  Best Local Similarity   54.3%;  Pred. No. 0.16;
  Matches   82;  Conservative    0;  Mismatches   69;  Indels    0;  Gaps    0;

Qy     199 TAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACG 258
           || | || ||| |  |||||| || | |  | || | |||   | | ||     ||||
Db    6009 TAAGCTAAATTTTTATCTTCTAATATATATACATATATATATATATATATTATATAAT 5950

Qy     259 AGTAGTTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATT 318
           |  | || ||   | |  || |||  ||  || ||||||||||  |  | |  | || | |
Db    5949 ATAAATTTTAATAAAACACAAATAATTAATATTACATATATCGTATGAATAATATACAGT 5890

Qy     319 AAATTAAAACTAAAAATGACTAAATTTCTAA 349
             ||| |   |   ||   | ||||   ||
Db    5889 TTCTTACAGAAATCTATATATTAATTAAAAA 5859


RESULT 9
US-08-599-252-96
; Sequence 96, Application US/08599252
; Patent No. 5705343
;  GENERAL INFORMATION:
;    APPLICANT:  DRAYNA, DENNIS T.
;    APPLICANT:  FEDER, JOHN N.
;    APPLICANT:  GNIRKE, ANDREAS
;    APPLICANT:  KIMMEL, BRUCE E.
;    APPLICANT:  THOMAS, WINSTON J.
;    APPLICANT:  WOLFF, ROGER K.
;    TITLE OF INVENTION:  METHOD TO DIAGNOSE HEREDITARY
;    TITLE OF INVENTION:  HEMOCHROMATOSIS
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Ave. N.W., Suite 5500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/599,252
;      FILING DATE:  09-FEB-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  9053-0001.21
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1000 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-599-252-96

  Query Match             3.1%;  Score 40.4;  DB 1;  Length 1000;
  Best Local Similarity   46.4%;  Pred. No. 0.072;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
           || | |  |  |||||       || |||| | ||   || |  ||||        ||||
Db     117 TTATTTATTATTTATTAAAAATAATATTTAATATTAAATATTATTTATTACTTATATTAT 176

Qy     268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
             ||   |  |   |  ||| ||     | || | || || ||  | |||||| |  ||
Db     177 TTATTATTATAATAATTTATAATATGTCACATTATAAAATATTATTTAATTAAAATTTAA 236

Qy     328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
           |     | ||  ||  || ||  ||  |||| |||  || | |  |  | ||
Db     237 TGTCACTCTATTATTTTAATATCATAAAATACATGAAAATACATTTTTCTGTAGAATCAC 296

Qy     388 CTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAA 447
           | |||  | |  |   | |    |  | ||||  |||     | | |||       ||
Db     297 GTTTTCCTCCTACTGTGAATAAAGACATAACTCACTAAGGGGAAGAATCTTGGCCCNAAG 356

Qy     448 ACAGTGCTTTGTACGACTATATA 470
           |||  |   |  | | | |
Db     357 TGTGTGATAAATCANANAANANA 379


RESULT 10
PCT-US96-06352-96
; Sequence 96, Application PC/TUS9606352
;  GENERAL INFORMATION:
;    APPLICANT:  DRAYNA, DENNIS T.
;    APPLICANT:  FEDER, JOHN N.
;    APPLICANT:  GNIRKE, ANDREAS
;    APPLICANT:  KIMMEL, BRUCE E.
;    APPLICANT:  THOMAS, WINSTON J.
;    APPLICANT:  WOLFF, ROGER K.
;    TITLE OF INVENTION:  METHOD TO DIAGNOSE HEREDITARY
;    TITLE OF INVENTION:  HEMOCHROMATOSIS
;    NUMBER OF SEQUENCES:  124
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  2000 Pennsylvania Ave. N.W., Suite 5500
;      CITY:  Washington
;      STATE:  DC
;      COUNTRY:  USA
;      ZIP:  20006-1888
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/06352
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/599,252
;      FILING DATE:  09-FEB-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  MURASHIGE, KATE H.
;      REGISTRATION NUMBER:  29,959
;      REFERENCE/DOCKET NUMBER:  9053-0001.21
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 887-1500
;      TELEFAX:  (202) 887-0763
;      TELEX:  90-4030
;  INFORMATION FOR SEQ ID NO:  96:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1000 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
PCT-US96-06352-96

  Query Match             3.1%;  Score 40.4;  DB 5;  Length 1000;
  Best Local Similarity   46.4%;  Pred. No. 0.072;
  Matches  122;  Conservative    0;  Mismatches  141;  Indels    0;  Gaps    0;

Qy     208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
```

```
Db    117 TTATTTATTATTTATTAAAAATAATATTTAATATTAAATATTATTTATTACTTATATTAT 176

Qy    268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
Db    177 TTATTATTATAATAATTTATAATATGTCACATTATAAAATATTATTTAATTAAAATTTAA 236

Qy    328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
Db    237 TGTCACTCTATTATTTTAATATCATAAAATACATGAAAATACATTTTTCTGTAGAATCAC 296

Qy    388 CTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAA 447
Db    297 GTTTTCCTCCTACTGTGAATAAAGACATAACTCACTAAGGGGAAGAATCTTGGCCCNAAG 356

Qy    448 ACAGTGCTTTGTACGACTATATA 470
Db    357 TGTGTGATAAATCANANAANANA 379
```

RESULT 11
PCT-US96-06583-96
; Sequence 96, Application PC/TUS9606583
; GENERAL INFORMATION:
; APPLICANT: DRAYNA, DENNIS T.
; APPLICANT: FEDER, JOHN N.
; APPLICANT: GNIRKE, ANDREAS
; APPLICANT: KIMMEL, BRUCE E.
; APPLICANT: THOMAS, WINSTON J.
; APPLICANT: WOLFF, ROGER K.
; TITLE OF INVENTION: METHOD TO DIAGNOSE HEREDITARY
; TITLE OF INVENTION: HEMOCHROMATOSIS
; NUMBER OF SEQUENCES: 124
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/06583
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,252
; FILING DATE: 09-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 9053-0001.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1000 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
PCT-US96-06583-96

Query Match 3.1%; Score 40.4; DB 5; Length 1000;
Best Local Similarity 46.4%; Pred. No. 0.072;
Matches 122; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy    208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
Db    117 TTATTTATTATTTATTAAAAATAATATTTAATATTAAATATTATTTATTACTTATATTAT 176

Qy    268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
Db    177 TTATTATTATAATAATTTATAATATGTCACATTATAAAATATTATTTAATTAAAATTTAA 236

Qy    328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
Db    237 TGTCACTCTATTATTTTAATATCATAAAATACATGAAAATACATTTTTCTGTAGAATCAC 296

Qy    388 CTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAA 447
Db    297 GTTTTCCTCCTACTGTGAATAAAGACATAACTCACTAAGGGGAAGAATCTTGGCCCNAAG 356

Qy    448 ACAGTGCTTTGTACGACTATATA 470
Db    357 TGTGTGATAAATCANANAANANA 379
```

RESULT 12
US-08-972-927-5/c
; Sequence 5, Application US/08972927
; Patent No. 6166290
; GENERAL INFORMATION:
; APPLICANT: Rea, Philip A
; APPLICANT: Lu, Yu-Ping
; APPLICANT: Li, Ze-Sheng
; TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: One Commerce Square, 2005 Market Street, 22nd
; STREET: Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: US
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,927
; FILING DATE: 18-NOV-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,040
; FILING DATE: 18-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/061,328
; FILING DATE: 08-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Doyle Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9596-12U2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10342 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-972-927-5

Query Match 3.1%; Score 40.4; DB 4; Length 10342;
Best Local Similarity 50.5%; Pred. No. 0.25;

```
  Matches   98;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy     331 AAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTA 390
           ||||  |  |||||| ||   | |   | ||| |||   ||| |  ||||| | | |  |
Db   10342 AAAAAAAGGAAATTTGTATTGCAATTTATATTTTAAGTGTTTTTTTAACAATTGTTCAAA 10283

Qy     391 TTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACA 450
           ||||  || | | ||| |  |||   ||  | |     | |  ||| |  |   || |
Db   10282 TTCTTTATAGATATATGTTTAATAGAACATTTTTGTGCAAAATGTCAATGACAAAACATG 10223

Qy     451 GTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAGTCGTTTG 510
           || || ||   ||   |||   | |||  || | |   |   |       |     ||| |
Db   10222 GTCCTGTGAGAGAAATTATGTCCAATGGTTGTCCATGTCCAGAAGTTGAAACCATGTTGG 10163

Qy     511 AAGATTGAACCTAT 524
           |  ||| ||  ||
Db   10162 CATGTTGCACTCAT 10149


RESULT 13
US-07-638-431-1
; Sequence 1, Application US/07638431
; Patent No. 5198535
;  GENERAL INFORMATION:
;    APPLICANT:  Hoffman, Stephen L.
;    APPLICANT:  Charoenvit, Yupin
;    APPLICANT:  Hedstrom, Richard
;    APPLICANT:  Khusmith, Srisin
;    APPLICANT:  Rogers IV, William O.
;    TITLE OF INVENTION:  Protective malaria sporozoite surface protein
;    TITLE OF INVENTION:  immunogen and gene
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  A. David Spevack
;      STREET:  NMRDC Building 1 T-12 National Naval
;      STREET:  Medical Center
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20814-5044
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/638,431
;      FILING DATE:  19910110
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, Avrom D.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-4033
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4673 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium yoelii
;      STRAIN:  17X(NL)
;      DEVELOPMENTAL STAGE:  erythrocytic stage
;      TISSUE TYPE:  Blood
;      CELL TYPE:  erythrocytic stage
;    IMMEDIATE SOURCE:
;      LIBRARY:  Py-lambdagt11-2-7 kb genomic expression
;      CLONE:  Py10.1111
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  718..3195
;      OTHER INFORMATION:
US-07-638-431-1

  Query Match             3.1%;  Score 40;  DB 1;  Length 4673;
  Best Local Similarity   52.4%;  Pred. No. 0.21;
  Matches   88;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
           | |||  | || ||      | |   | |   |||| || ||||   | |   |  ||
Db    3487 TTCATATTTTTTTTTAACGATTTTTTTTTAACTATTTTTTTTTAACTAATTGTCTCGTTA 3546

Qy     264 TTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATT 323
           | ||| ||| ||  |  || |   ||  |  ||      |  ||| |||| |||  ||
Db    3547 TAATATATATATTTATTTATACTCCAATATTTAATGGTTACAATTATTCTTAATATAAAA 3606

Qy     324 AAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTT 371
           ||||  |||||  |  |||  |  ||    ||  ||||| | | ||||
Db    3607 AAAAAAAAAAAAAAAAAAAACTTAAAAGTTAATAACATTTTTAGGTTT 3654


RESULT 14
PCT-US92-00018-1
; Sequence 1, Application PC/TUS9200018
;  GENERAL INFORMATION:
;    APPLICANT:  Hoffman, Stephen L.
;    APPLICANT:  Charoenvit, Yupin
;    APPLICANT:  Hedstrom, Richard
;    APPLICANT:  Khusmith, Srisin
;    APPLICANT:  Rogers IV, William O.
;    TITLE OF INVENTION:  Protective malaria sporozoite surface protein
;    TITLE OF INVENTION:  immunogen and gene encoding
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  A. David Spevack
;      STREET:  NMRDC Building 1 T-12 National Naval
;      STREET:  Medical Center
;      CITY:  Bethesda
;      STATE:  MD
;      COUNTRY:  USA
;      ZIP:  20814-5044
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US92/00018
;      FILING DATE:  19920103
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Spevack, Avram D.
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 295-6759
;      TELEFAX:  (301) 295-4033
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4673 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium yoelii
;      STRAIN:  17X(NL)
;      DEVELOPMENTAL STAGE:  erythrocytic stage
;      TISSUE TYPE:  Blood
;      CELL TYPE:  erythrocytic stage
;    IMMEDIATE SOURCE:
;      LIBRARY:  Py-lambdagt11-2-7 kb genomic expression
```

```
;      CLONE:  Py10.1111
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  718..3195
;      OTHER INFORMATION:
PCT-US92-00018-1

  Query Match             3.1%;  Score 40;  DB 5;  Length 4673;
  Best Local Similarity   52.4%;  Pred. No. 0.21;
  Matches   88;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
           | |||  | || ||       | |   | |   |||| || ||||   | |   |  ||
Db    3487 TTCATATTTTTTTTTAACGATTTTTTTTTAACTATTTTTTTTTAACTAATTGTCTCGTTA 3546

Qy     264 TTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATT 323
           | ||| ||| ||  |  || |   ||  |  ||      |  ||| |||| ||| ||
Db    3547 TAATATATATATTTATTTATACTCCAATATTTAATGGTTACAATTATTCTTAATATAAAA 3606

Qy     324 AAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTT 371
           ||||  |||||  |  |||  |  ||    ||  ||||| | | ||||
Db    3607 AAAAAAAAAAAAAAAAAAAACTTAAAAGTTAATAACATTTTTAGGTTT 3654


RESULT 15
US-09-004-838-109/c
; Sequence 109, Application US/09004838
; Patent No. 6350933
;  GENERAL INFORMATION:
;    APPLICANT:  Michelmore, Richard W.
;    APPLICANT:  Shen, Kathy
;    APPLICANT:  Meyers, Blake
;    TITLE OF INVENTION:  Procedures and Materials for
;    TITLE OF INVENTION:  Conferring Pest Resistance in Plants
;    NUMBER OF SEQUENCES:  140
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew LLP
;      STREET:  Two Embarcadero Center, Eighth Floor
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94111-3834
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/004,838
;      FILING DATE:  09-JAN-1998
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/781,734
;      FILING DATE:  10-JAN-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Einhorn, Gregory P.
;      REGISTRATION NUMBER:  38,440
;      REFERENCE/DOCKET NUMBER:  023070-078810US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 576-0200
;      TELEFAX:  (415) 576-0300
;  INFORMATION FOR SEQ ID NO:  109:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5829 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA
;    FEATURE:
;      NAME/KEY:  -
;      LOCATION:  1..5829
;      OTHER INFORMATION:  /note= "RG2K"
US-09-004-838-109

  Query Match             3.1%;  Score 40;  DB 4;  Length 5829;
  Best Local Similarity   45.9%;  Pred. No. 0.24;
  Matches  136;  Conservative    0;  Mismatches  160;  Indels    0;  Gaps    0;

Qy     209 TCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATA 268
           | | || | ||||  | | |  | || |  ||  | |||   | | |     |  || ||
Db    4460 TATATTTTGCTTAAATCTTCCAAATTCTTATGAGGATTAGCCAATCTTGTCTTTTTTCTA 4401

Qy     269 CATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAAC 328
           | | ||    ||   |   |||  |  ||  || |  |  |   ||||||  |   | |
Db    4400 TAAAAATAATAATGAGTGATATCTTAACTAAAACTAGTTAATAAAAAATTAGTTACGATC 4341

Qy     329 TAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACC 388
           || || ||||      | |   | || |   |   | ||  ||      |  |  | ||
Db    4340 TATAACTGACATTGGCTATTTAAACATCTTTAAGCTTATTGTTATAAAATGATATATAAT 4281

Qy     389 TATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAA 448
           |    |     || |  |    |||| | |  ||||| ||  | |  |   ||  | |
Db    4280 AAAGAAAATCATTTATTGTAGACATTTTAATAAATAAAAAAGGTTGGTTTACATTTATCA 4221

Qy     449 CAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAGT 504
           |  | || |   | |||||     || | | || |||  |    | || || ||
Db    4220 ATGACCATTTTTTAATTATATTATGCATTTATTGCAACATTAGTTCAAAAAATGGT 4165


Search completed: November  7, 2002, 10:48:43
Job time : 86.9851 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 08:30:58 ; Search time 1237.03 Seconds
(without alignments)
14282.232 Million cell updates/sec

Title: US-09-905-558C-16
Perfect score: 1309
Sequence: 1 cccatcgctgctttgtctac.....caagggaaagtgatcccatg 1309

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	149.2	11.4	374	12	BH255115	BH255115 RXL1BAM00
c	2	123.8	9.5	223	12	AZ921021	AZ921021 1006023E0
c	3	107.6	8.2	496	9	AW928053	AW928053 945007D02
c	4	58.2	4.4	1101	12	CNS0106X	AL098595 Drosophil
c	5	57.8	4.4	500	9	AU087781	AU087781 AU087781
c	6	57.4	4.4	935	12	CNS006XK	AL066051 Drosophil
	7	56.6	4.3	534	12	BH255739	BH255739 LDH5BAM00
	8	55.6	4.2	1058	12	CNS014D0	AL103998 Drosophil
c	9	55.4	4.2	745	12	AZ046300	AZ046300 nbeb0091L
c	10	55.2	4.2	776	12	AG130889	AG130889 Pan trogl
c	11	54.2	4.1	818	10	BI952390	BI952390 HVSMEm000
	12	54	4.1	798	12	AQ251130	AQ251130 F26H6-Sp6
	13	54	4.1	1101	12	CNS00FVE	AL071298 Drosophil
	14	53.4	4.1	1101	12	CNS008WC	AL052719 Drosophil
	15	53.2	4.1	945	12	CNS04D0K	AL285149 Tetraodon
c	16	53	4.0	910	12	CNS006ON	AL065629 Drosophil
c	17	53	4.0	914	12	AZ674804	AZ674804 ENTKN64TR
c	18	53	4.0	928	12	CNS00DKY	AL071865 Drosophil
	19	53	4.0	936	10	BG852371	BG852371 1024034A0
	20	53	4.0	1101	12	CNS0039G	AL063921 Drosophil
	21	52.8	4.0	652	12	BH255629	BH255629 LDH5BAM00
	22	52.8	4.0	928	12	CNS00DKY	AL071865 Drosophil
c	23	52.6	4.0	848	10	BI950718	BI950718 HVSME1002
c	24	52.6	4.0	1092	12	CNS020K7	AL175696 Tetraodon
c	25	52.6	4.0	1101	12	CNS017KE	AL108152 Drosophil
c	26	52.6	4.0	1101	12	CNS017ZQ	AL108704 Drosophil
c	27	52.4	4.0	479	10	BM163977	BM163977 EST566500
	28	52.4	4.0	1101	12	CNS00G7F	AL071731 Drosophil
	29	52.2	4.0	914	12	CNS00ZJY	AL097768 Drosophil
c	30	52.2	4.0	1101	12	CNS00H18	AL072804 Drosophil
c	31	52	4.0	861	10	BI953364	BI953364 HVSMEm001
	32	52	4.0	942	9	AL536132	AL536132 AL536132
c	33	52	4.0	1059	12	CNS0155J	AL105025 Drosophil
	34	52	4.0	1101	12	CNS017EJ	AL107941 Drosophil
c	35	51.8	4.0	802	12	CNS011T0	AL100686 Drosophil
c	36	51.8	4.0	1043	12	CNS0145P	AL103735 Drosophil
c	37	51.8	4.0	1101	12	CNS00Z3I	AL097176 Drosophil
c	38	51.6	3.9	776	12	AG060484	AG060484 Pan trogl
c	39	51.6	3.9	969	12	CNS03PO0	AL254889 Tetraodon
	40	51.4	3.9	553	9	AI657313	AI657313 486093C05
c	41	51.4	3.9	563	10	BG004199	BG004199 QV4-GN012
	42	51.4	3.9	925	10	BG441241	BG441241 GA__Ea001
	43	51.4	3.9	1101	12	CNS016LI	AL106896 Drosophil
c	44	51.2	3.9	342	10	BM168596	BM168596 EST571119
c	45	51.2	3.9	358	10	BM168541	BM168541 EST571064

ALIGNMENTS

RESULT 1
BH255115/c
LOCUS BH255115 374 bp DNA linear GSS 29-NOV-2001
DEFINITION RXL1BAM0001A01r Zea mays L. XL1 methyl filtration maize root genomic shotgun library Zea mays genomic clone RXL1BAM0001A01r, DNA sequence.
ACCESSION BH255115
VERSION BH255115.1 GI:17150008
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 374)
AUTHORS Kim,S.W., Yu,Y., Lee,M.C., Yang,T.J., Main,D., Henry,D., Oates,R. and Wing,R.A.
TITLE Genomic shotgun library from maize
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACCCTCACTAAAGGG
Class: shotgun
High quality sequence stop: 374.
FEATURES Location/Qualifiers
source 1. .374
/organism="Zea mays"
/strain="B73"
/db_xref="taxon:4577"
/clone="RXL1BAM0001A01r"
/clone_lib="Zea mays L. XL1 methyl filtration maize root genomic shotgun library"
/tissue_type="Root"
/lab_host="XL1Blue"
/note="Vector: pCUGIblu-1; Site_1: Sau3A1;

Methyl-filtration library, Nuclei DNA was digested with Sau3A1, size fractionated and transformed to E.Coli.XL1Blue."
BASE COUNT 125 a 64 c 70 g 115 t
ORIGIN

Query Match 11.4%; Score 149.2; DB 12; Length 374;
Best Local Similarity 79.9%; Pred. No. 5.2e-21;
Matches 214; Conservative 0; Mismatches 48; Indels 6; Gaps 3;

```
Qy     78 CCGTTCGAGTGACTGCATGGCGTACATCTTTCTGCATCGACTTTGTACGGCTACATCGAA 137
           || |||||| |||| ||  |||||||| |  |||||||||||    ||   ||| ||||||
Db    356 CCATTCGAGGGACTACACTGCGTACATTTGCCTGCATCGACTGCATA---CTAAATCGAA 300

Qy    138 CATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAAGCATCGGTTACTCC 197
           |||| |||||||||||| |||||||||||  ||   ||| ||| |||||||||  || |
Db    299 CATACACACGAGATGTCTTGTGTGAATAGAGCCAACGATGTCTTGAGCATCGGTCCCTTC 240

Qy    198 GTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATAC 257
           | |||||||||| |||||||  |  |||||||  ||||| |||||||||| | |||| |
Db    239 GCAGGGTACATTATGTTCTTAGTGATTGTGCAT-GTTTTACTGTTGTTTACCGCTTATGC 181

Qy    258 GAGTAGTTATACATACATGCAC--ATACATATCATCACATATATCACAATATTTTTCTAA 315
           |||||||  |||| ||||||||  |||||| |||||||||| |||||| |  ||| ||
Db    180 GAGTAGTAGTACACACATGCACATATACATGTCATCACATACATCACACTGGTTTCCTGG 121

Qy    316 ATTAAATTAAAACTAAAAATGACTAAAT 343
           |||||||||||||| | |||||||| ||
Db    120 ATTAAATTAAAACTGATAATGACTATAT 93
```

RESULT 2
AZ921021/c
LOCUS AZ921021 223 bp DNA linear GSS 17-DEC-2001
DEFINITION 1006023E04.2EL_y1 1006 - RescueMu Grid G Zea mays genomic, DNA sequence.
ACCESSION AZ921021
VERSION AZ921021.1 GI:13392238
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 223)
AUTHORS Walbot,V.
TITLE Maize genomic sequences found using engineered RescueMu transposon
JOURNAL Unpublished (2001)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Possible ligation site of ends cut by 2 different endonucleases.
Reverse complemented post-ligation sequence from source sequence.
Plate: 1006023 row: 39
Class: transposon-tagged.
FEATURES Location/Qualifiers
source 1. .223
/organism="Zea mays"
/cultivar="mixed background W23/A188/B73"
/db_xref="taxon:4577"
/clone_lib="1006 - RescueMu Grid G"
/tissue_type="leaf"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: leaf; Vector: RescueMu (engineered from pBlueScript backbone); Site_1: BamHI; Site_2: BglII; RescueMu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on RescueMu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'RescueMu.' Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."
BASE COUNT 73 a 36 c 46 g 68 t
ORIGIN

Query Match 9.5%; Score 123.8; DB 12; Length 223;
Best Local Similarity 82.4%; Pred. No. 9.9e-16;
Matches 154; Conservative 0; Mismatches 32; Indels 1; Gaps 1;

```
Qy    166 GAGTCACTAATGCCTTAAGCATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTG 225
           ||  ||||||||| || |||||||||   |||||||||||||||||||||||||  || |||
Db    223 GATCCACTAATGCTTTGAGCATCGGTCCATCCGTAGGGTACATTCTGTTCTTAGTAATTG 164

Qy    226 TGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACAT 285
           |||||  ||||| || |||||||| |||| |||||| | ||||| | ||||||||  |||
Db    163 TGCAT-GTTTTAATGCTGTTTACTGCTTATGCGAGTATTCATACACATATGCACATGCAT 105

Qy    286 ATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTT 345
           | ||||||| |||||| || ||||||  |||||||||||| |||||||  |||| | |
Db    104 GCCGTCACATACATCACACTAATTTTCTGGATTAAATTAAAACAAAAAATGCCTAATTCT 45

Qy    346 CTAACAC 352
           |||||||
Db     44 CTAACAC 38
```

RESULT 3
AW928053/c
LOCUS AW928053 496 bp mRNA linear EST 30-MAY-2000
DEFINITION 945007D02.Y1 945 - Mixed adult tissues from Walbot lab, same as 707 (SK) Zea mays cDNA, mRNA sequence.
ACCESSION AW928053
VERSION AW928053.1 GI:8103414
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 496)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 945007 row: D column: 02.
FEATURES Location/Qualifiers
source 1. .496
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="945 - Mixed adult tissues from Walbot lab, same as 707 (SK)"
/tissue_type="tassel, kernal, silk, husk, root, leaf"
/dev_stage="fully-grown"
/lab_host="DH10B"
/note="Organ: tassel, kernal, silk, husk, root, leaf; Vector: pGAD10; Site_1: EcoRI; cDNA library from fully differentiated maize tissues from an active Mutator plant. Tissue ratio is 4:2:1:1:1:1 (tassel, kernel, silk, husk, root, leaf). Unidirectionally cloned. New library number given to library 707 for additional sequencing."

```
BASE COUNT      203 a     62 c    107 g    124 t
ORIGIN

  Query Match             8.2%;  Score 107.6;  DB 9;  Length 496;
  Best Local Similarity   78.0%;  Pred. No. 2.4e-12;
  Matches  142;  Conservative    0;  Mismatches   39;  Indels    1;  Gaps    1;

Qy    208 TTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTAT 267
          || | | |||| ||||||||||  ||| |||| |||||||| |||| ||||||||||
Db    429 TTTTTTCCTTCGTATTTGTGCAT-GTTTCATTGCTGTTTACTGCTTATGCGAGTAGTTAT 371

Qy    268 ACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAA 327
          || |||||||||||||| |||||||||||||| || | |||||||  ||||||||||||
Db    370 GCACACATGCACATACATGTCATCACATATATCGCACTGTTTTTCTGGATTAAATTAAAA 311

Qy    328 CTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTAC 387
          ||||||||| |||| |||||||||   || | | | | ||  | |    ||| |  |||
Db    310 CTAAAAATGCCTAACTTTCTAACAAATACTAAAGTTTGTTGCGTCCAAACACATCAATAC 251

Qy    388 CT 389
          |
Db    250 AT 249


RESULT 4
CNS0106X/c
LOCUS       CNS0106X                1101 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC
            BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL098595
VERSION     AL098595.1  GI:5610206
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03K20"
                     /note="end : T7"
BASE COUNT      258 a    107 c     60 g    175 t    501 others
ORIGIN

  Query Match             4.4%;  Score 58.2;  DB 12;  Length 1101;
  Best Local Similarity   19.2%;  Pred. No. 0.05;
  Matches   87;  Conservative  182;  Mismatches  184;  Indels    0;  Gaps    0;

Qy    209 TCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATA 268
          |:| ||::|:|   |  :  :|:|:|||:::.|   |  :|  :|:|:    |: :|: :
Db   1027 TMTMTTHHTMTMTMTMTHTTMTMTHTTHHHTTMTMTMTMTMTHTHTMTMTWTMTHTHMM 968

Qy    269 CATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAAC 328
          :::::::| :::::::::|::: ::::: : :|: ::::: ::: :::::|:: :|  : ::

Db    967 HMHMHMTMHMMHHHMMTMHMTYHMMWMMMHTMTHMMHHMHHHHMHHMHMTHMTMTMTMTHM 908

Qy    329 TAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACC 388
          :::::::  :|::: |  |:: :   :  ::: | |::|  |:|:|::::|  :: |: :
Db    907 HMMMMMMMTMTMHMMTMATMMTMATTMTHHHHMTMTMMTTMTHTMTMHMHCTYYYMTMTH 848

Qy    389 TATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAA 448
            |:   : || : :|:  :|  ::  ::: :::| ::: :|  : |:| : :|  :  :
Db    847 CTTHTMTMTTTMHMMTMMMHCWTMHMMMMMMYHMTYMMMCHCYCHHTHTTHTMTMTHTTM 788

Qy    449 CAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAGTCGTT 508
          : |  :|: | :  :|:|   : :  : :      : : : :|::  |::| |: |   :
Db    787 THMTHTHTHHTTMTTMTMTYMYMHMMTHWYYMHYTHTMTMTMHMTCMHTTTHTTATAW 728

Qy    509 TGAAGATTGAACCTATATATCGGTACGGTTAATCCGTCTATGTACGTGGGCATGACGAAC 568
            |:  :|   : |:|:|:|:  : :   ||:|   |:|:|  ::      :: :  ::
Db    727 ATAHWTYTMTYMATMTMTMTMTYTHMCATAMTTTMTMTMTTMMMTMTMATMHWMWTHMT 668

Qy    569 ACCCGTGATAACGAAGGATTAACGTGCACAATCATAAATCCAAAGTAGGAGCGGTGCATG 628
          |::: : : :|   |   |  |:    |::  :|| |: :  |:  | :| :  | :|
Db    667 AHMHMYTMMMATMTATATTAYAHMAHTAYMTAMATTAWAYATAMACATKATMAMTMHAAC 608

Qy    629 ATGAGAATCGCTCTCAGTACTCGACATAATGAA 661
          |     :: :   : :  |:: || |:::  |:
Db    607 AMACMTMWAKAAAWAMAAAMDAGAGADMMACAR 575


RESULT 5
AU087781/c
LOCUS       AU087781                500 bp    mRNA    linear   EST 27-JAN-2001
DEFINITION  AU087781 Sugano Malaria cDNA library Plasmodium falciparum cDNA
            clone XPFn5810, mRNA sequence.
ACCESSION   AU087781
VERSION     AU087781.1  GI:12389922
KEYWORDS    EST.
SOURCE      malaria parasite P. falciparum.
  ORGANISM  Plasmodium falciparum
            Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE   1  (bases 1 to 500)
  AUTHORS   Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.
  TITLE     FULL-malaria: a database for a full-length enriched cDNA library
            from human malaria parasite, Plasmodium falciparum
  JOURNAL   Nucleic Acids Res. 29 (1), 70-71 (2001)
  MEDLINE   20574754
COMMENT     Contact: Junichi Watanabe
            Institute of Medical Science
            The University of Tokyo, Department of Parasitology
            4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan
            Tel: 81-3-5449-5378
            Fax: 81-3-5449-5410
            Email: jwatanab@manage.ims.u-tokyo.ac.jp
            Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano
            ,S. Construction and characterization of a full length-enriched and
            a 5'-end-enriched cDNA library Gene 200 (1-2), 149-156 (1997).
FEATURES             Location/Qualifiers
     source          1. .500
                     /organism="Plasmodium falciparum"
                     /strain="3D7"
                     /db_xref="taxon:5833"
                     /clone="XPFn5810"
                     /clone_lib="Sugano Malaria cDNA library"
                     /dev_stage="erythrocytic stage"
BASE COUNT      228 a     25 c     33 g    203 t     11 others
ORIGIN

  Query Match             4.4%;  Score 57.8;  DB 9;  Length 500;
  Best Local Similarity   54.2%;  Pred. No. 0.059;
  Matches  136;  Conservative    0;  Mismatches  114;  Indels    1;  Gaps    1;

Qy    220 TATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCAC 279
          ||| | |  |||| |||| |  | |||| | |||  |   ||| ||| | ||| || |||
Db    274 TATATATATATATATTTACTAATTTTTATTTATTTAAAATGTAATTAAAAATATATACAC 215
```

```
Qy     280 ATACATATCATCACATATATCACAATATTTTTTCTAAATTAAATTAAAACTAAAAATGACT 339
           ||| ||||    | ||||||    |||| | | || ||  | |||    ||    |   |
Db     214 ATATATATATATATATATATATATATATATATATATATTTATGTGTATTTTTTTTT 155

Qy     340 AAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATT 399
           | | | |||    ||    | | |||||| |  | | |  | | ||  ||||  | |
Db     154 TATTAT-TAAATTTAATTTTAATCTAATGTGTGTTTCTAATAATATAAATATTANATAAA 96

Qy     400 GTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGT 459
           || | |||  ||  |    | | |   || |  | |||   |  ||||  | | ||   |
Db      95 TTTTTTTTTATAAAATNCGACGAAAGTGAAAAATGACTATTGTAAAAAATATTTCTGGAT 36

Qy     460 ACGACTATATA 470
           ||    ||| |
Db      35 ACATAAATACA 25
```

RESULT 6
CNS006XK/c
LOCUS CNS006XK 935 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL066051
VERSION AL066051.1 GI:4945019
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 935)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .935
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR14N09"
/note="end : T7"
BASE COUNT 257 a 170 c 162 g 96 t 250 others
ORIGIN

Query Match 4.4%; Score 57.4; DB 12; Length 935;
Best Local Similarity 30.2%; Pred. No. 0.073;
Matches 115; Conservative 86; Mismatches 180; Indels 0; Gaps 0;

```
Qy     803 CGAGGTGGTGCCCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCA 862
           :| :| |:   : :| :|| : : || |:  | :: : ::  ||| : :|  :||| :
Db     921 SGSSGCGSGCGSGSGGSCCCCSGSCCCSCSCGCCSSCSCSSSSCCCSSCSCGCSCGCCSC 862

Qy     863 CCAGCCATGCCGTGTCGCGGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTC 922
             :||  ||   | ::  |  |:|  |  || :|  :  ||: : | ||   :|:||
Db     861 GGSSCCSCGCGCSGSSSGCGCCCSGGSCGCCCGSCGGSCSCCSSSCCGCGCCGSCSCGCG 802

Qy     923 CCGGCCGTGCCCGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCC 982
           |||||:|  |: | | :|   |: ||  ::|  |: ::  || ||:|   : : |  :|
Db     801 CCGGCSGGCCSGGCGSSGCGGCSGGCCGSSCCGCSGKSKCGCGCGSGGSCSSSSGSGSCG 742

Qy     983 TGGACACCCGACACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCA 1042
            :| : |::  : |:|   ::|: |   |      |   |:           : :| |
Db     741 GSGGSGCSSCSSGCSCGCGSSGSGCGSCGCCCSCGCSCGCSCGCSSGCGSCGCSGSCGCG 682

Qy    1043 CGGCCGACTGCGCTGCGCCGCCCGGATATATATACCCATCGTTATCGATCGATCGATCGC 1102
           | |||:    |:|  |:|::::||:          |::  ::    ||  |:  ||  |:|
Db     681 CCGCCSCGCCCSCCCCSCSSSSCGSSCSSCCSSSCSSCCSSSGCCCGCCCSCGCGCGCSC 622

Qy    1103 GTCACTCACGGGTAGCTCATGGTCGAGCGTAGCATGCAGGAACTTATTTGCCGTGCGCTC 1162
           | | | | | || ||| || |  :|:  :  : | :::::  :      : || |||: |
Db     621 GCCGCSCCCCGGCAGCSCASGCSSGMAGVAGSGASRSVVSGGSSSSGSASGCGCGCGMGC 562

Qy    1163 CCAGGTCTCCGCTCGCGTGCC 1183
           :|:: : |:  :||  |:
Db     561 RAGSGKMGSAGSSGRCGGACS 541

RESULT 7
BH255739
LOCUS BH255739 534 bp DNA linear GSS 29-NOV-2001
DEFINITION LDH5BAM0003C03f Zea mays L. DH5alpha methy filtration maize leaf genomic shotgun library Zea mays genomic clone LDH5BAM0003C03f, DNA sequence.
ACCESSION BH255739
VERSION BH255739.1 GI:17150632
KEYWORDS GSS.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 534)
AUTHORS Kim,S.W., Yu,Y., Lee,M.C., Yang,T.J., Main,D., Henry,D., Oates,R. and Wing,R.A.
TITLE Genomic shotgun library from maize
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
Class: shotgun
High quality sequence stop: 531.
FEATURES Location/Qualifiers
source 1. .534
/organism="Zea mays"
/strain="B73"
/db_xref="taxon:4577"
/clone="LDH5BAM0003C03f"
/clone_lib="Zea mays L. DH5alpha methy filtration maize leaf genomic shotgun library"
/tissue_type="Leaf"
/lab_host="DH5alpha"
/note="Vector: pCUGIblu-1; Site_1: Sau3A1; Site_2: Sau3A1; Methyl-filtration library, Nuclei DNA was digested with Sau3A1, size fractionated and transformed to E.Coli.DH5alpha."
BASE COUNT 174 a 77 c 78 g 205 t
ORIGIN

Query Match 4.3%; Score 56.6; DB 12; Length 534;
Best Local Similarity 58.2%; Pred. No. 0.1;
Matches 139; Conservative 0; Mismatches 94; Indels 6; Gaps 2;

```
Qy   204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
```
```

```
Db    236 TTCAGTTTTTTCTACTATCATGAGAATAATATATTGTATGTTTGTTACTATTTAGAGTAG 295

Qy    264 TTATACATACATGCACATACATATCATCACATATATCACAATAT-TTTTCTAAATTAAAT 322

Db    296 T-----ACACATGCACGTAAATGTCATCACATATTTTATGCAATGTTTTTTGGATTAAAT 350

Qy    323 TAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAAC 382

Db    351 TAAGCAATAATTTGCCTAAATTTCTAACACCAACTACATAAGATTTTTCATGTTTAATAC 410

Qy    383 TTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAA 441

Db    411 TTGTCAAAAGCTAGTTAATTAGGAGGGTTATTATAGTGTATAGGAAATATGGAGAATAA 469


RESULT 8
CNS014D0
LOCUS       CNS014D0                1058 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN11I17 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103998
VERSION     AL103998.1  GI:5615609
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1058)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1. .1058
                     /organism="Drosophila melanogaster"
                     /plasmid="pBeloBAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN11I17"
                     /note="end : SP6"
BASE COUNT      121 a    175 c     93 g    396 t    273 others
ORIGIN

  Query Match             4.2%;  Score 55.6;  DB 12;  Length 1058;
  Best Local Similarity   32.9%;  Pred. No. 0.17;
  Matches   78;  Conservative   65;  Mismatches   94;  Indels    0;  Gaps    0;

Qy    186 ATCGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTT 245

Db    700 ATAGSCWATTCWSASTRATASWMTSTSWTWBTBWVKRSTRATKAKTBATCTVTTWVTATA 759

Qy    246 TACTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACAAT 305

Db    760 HSMTSWWYTTACCMTSMCTSWTSTACCACWCYCTACCCAWTATATCTMMTATATHCACTW 819

Qy    306 ATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTA 365

Db    820 ACATTSCYTMAHTAYAHTTCACTWAANAAYTATTSTAWATYMTTSTSCTTSWVSAAAAYA 879

Qy    366 ATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTA 422

Db    880 TASWTTBTACSMTTACTSAYASYTHTAMTATATYATAAWWHYTMWATTHTYAHCATA 936


RESULT 9
AZ046300/c
LOCUS       AZ046300                745 bp    DNA     linear   GSS 16-MAR-2000
DEFINITION  nbeb0091L15r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
            clone OSJNBb0091L15r, DNA sequence.
ACCESSION   AZ046300
VERSION     AZ046300.1  GI:7250069
KEYWORDS    GSS.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   Wing,R.A. and Dean,R.A.
  TITLE     A BAC End Sequencing Framework to Sequence the Rice Genome
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: GCTATTTAGGTGACACTATAG
            Class: BAC ends
            High quality sequence stop: 213.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Oryza sativa"
                     /strain="Japonica"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:4530"
                     /clone="OSJNBb0091L15r"
                     /clone_lib="CUGI Rice BAC Library (EcoRI)"
                     /tissue_type="Leaf"
                     /lab_host="E. coli DH10B"
                     /note="Vector: pBACIndigo; Site_1: EcoRI; Site_2: EcoRI;
                     Rice is the most important food crop in the world. Half of
                     the world population, especially those inhabiting highly
                     populated areas of the humid tropics and subtropics, rely
                     on rice as their primary source of carbohydrate.
                     Monocotyledonous rice is a diploid plant (2n=24) with a
                     haploid genome equivalent of 431 Mbp (Arumuganathan and
                     Earle, 1991). The relatively small genome of rice, three
                     times larger than that of Arabidopsis, makes it suitable
                     for genomic studies. In order to facilitate positional
                     cloning, physical mapping and genome sequencing of rice,
                     we have constructed a BAC library from Oryza sativa,
                     Nipponbare variety using EcoRI as the cloning enzyme. The
                     library contains 55,296 clones with an average insert size
                     of 121 Kb providing approximatley 15 haploid genome
                     equivalents. The deep coverage allows the isolation a
                     particular sequence with a probability of 99.9 %. Three
                     high density filters, each containing 18,432 clones
                     (doubly spotted), represent the whole library for colony
                     screening and can be requested from the Clemson University
                     BAC/EST Resource Center (www.genome.clemson.edu)."
BASE COUNT      109 a     37 c    382 g     99 t    118 others
ORIGIN

  Query Match             4.2%;  Score 55.4;  DB 12;  Length 745;
  Best Local Similarity   36.7%;  Pred. No. 0.19;
  Matches  146;  Conservative    0;  Mismatches  252;  Indels    0;  Gaps    0;

Qy    702 CCGCCTTTATTTCAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACC 761

Db    636 CCNCNNNNNCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCNNNNNCNNCCCCCCC 577

Qy    762 CCAACCTCTCACCGAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGCC 821
```

```
          ||  || | | ||      | | |     |  |    |   |          ||||   ||
Db    576 CCNCCCCCCCCCCCCNNNNCNCCCNCCCCCCCCCCNCCCCCCNCNCCCCCCCCCCCCCCC 517

Qy    822 CACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGCG 881
          | | |  | |  || | |   | ||| |  |  || | |  |  ||   ||    | |
Db    516 CCCCCNNCCCCCCCCCCCNCNCCNCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCC 457

Qy    882 GAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCGTGTCGG 941
                |  || |  ||  ||| | | |    ||||  |||  ||   |||    |
Db    456 NNCNCNNCCNCCCNCNNCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCNCCCCCCCCC 397

Qy    942 TCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTGT 1001
            ||  |   |  |  |    |  ||| | |        ||||    | |||  | |||
Db    396 CCCCCCNNCCNCCCCCCCCCCCCCCCCCNCNNNNCNCCCCCCNCCCCCCCCNNCCCCCCC 337

Qy   1002 CGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCACGGCCGACTGCGCTGCGCC 1061
            |                  ||              | ||     ||          | |
Db    336 NNNCNNNNNNNNNNNNNNNCCNNNNNNNNNCNCNCCNNNNCCCNNNNNNCNNNCNCN 277

Qy   1062 GCCCGGATATATATACCCATCGTTATCGATCGATCGAT 1099
              |||| |||||||    |   |||  ||  ||  ||
Db    276 NTNCGGACATATATATATGTATATATATATATATATAT 239


RESULT 10
AG130889/c
LOCUS       AG130889               776 bp    DNA     linear   GSS 04-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-142M17.F, genomic survey sequence.
ACCESSION   AG130889
VERSION     AG130889.1  GI:16660054
KEYWORDS    GSS; GSS (genome survey sequence).
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-142M17.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (sites)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 776)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: -21M13
             LIBRARY
                 Vector    : pKS145
                 R.Site 1  : SacI
                 R.Site 2  : SacI.
FEATURES             Location/Qualifiers
     source          1. .776
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-142M17.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT       22 a     63 c    638 g     24 t     29 others
ORIGIN

  Query Match             4.2%;  Score 55.2;  DB 12;  Length 776;
  Best Local Similarity   47.0%;  Pred. No. 0.21;
  Matches  171;  Conservative    0;  Mismatches  193;  Indels    0;  Gaps    0;

Qy    702 CCGCCTTTATTTCAGGCTCGCTGGCCCCACGGGCGTGCTGCGTGCACGAAGGGCACTACC 761
          |||||       |  || | |   |||| || ||   |   | || |     || | ||
Db    730 CCGCCCCCCCCCCCCGCCCCCGCCCCCCCCGCGCCCTCCCGGCGCCCCCGCCGCCCCCCC 671

Qy    762 CCAACCTCTCACCGAAAACCGCGCTGGATCGGCAAATCAAACGAGGTGGTGCCCCGTGCC 821
          ||  ||   | ||      || | | |   |       |   | |      ||||  ||
Db    670 CCCCCCCGCCCCCCCCCGCCCCCCCGCCACCCGCGCCCCCCCCCGCCCCCCCCCCGCCC 611

Qy    822 CACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCCGTGTCGCG 881
          | | || |  || | | ||  ||| |  |  || | | ||  ||   ||    ||||
Db    610 TCCCCGCCGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGCCCTCCCCCCCCGCCCGCG 551

Qy    882 GAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCCCGTGTCGG 941
            |  | |  || |||||  ||| | | |    || |  | ||   |  |||    |
Db    550 CCCCCCCCCCCCCCCCCCCCCCCCCCCGCGCCGCCCCGCCCCGCGCGGGCCCCCCCCCCC 491

Qy    942 TCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGACACCCTGT 1001
            || | | |  | |   ||  ||| | |  |      ||||    | |||  | |||
Db    490 CCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCGCCCCCCCCGCCCCCCCCCCCC 431

Qy   1002 CGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCACGGCCGACTGCGCTGCGCC 1061
          |  |||||| |||   |  ||| |||    | |  ||||| |  ||  |  ||  | |||
Db    430 CCCCCCTTTTTTTCCCCCCCCCCAAACAAAACCCCCCCCCCCCCCCCCCCCCGGGGGGCC 371

Qy   1062 GCCC 1065
            |||
Db    370 CCCC 367


RESULT 11
BI952390/c
LOCUS       BI952390               818 bp    mRNA    linear   EST 19-OCT-2001
DEFINITION  HVSMEm0006C23f Hordeum vulgare green seedling EST library
            HVcDNA0014 (Blumeria infected) Hordeum vulgare cDNA clone
            HVSMEm0006C23f, mRNA sequence.
ACCESSION   BI952390
VERSION     BI952390.1  GI:16296163
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
            Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons
            ,J., Oates,R. and Main,D.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics: Blumeria infected Morex (compatible) seedling
            cDNA library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Total hq bases = 224
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence stop: 808.
FEATURES             Location/Qualifiers
     source          1. .818
                     /organism="Hordeum vulgare"
                     /cultivar="Morex"
                     /db_xref="taxon:4513"
                     /clone="HVSMEm0006C23f"
                     /clone_lib="Hordeum vulgare green seedling EST library
                     HVcDNA0014 (Blumeria infected)"
                     /tissue_type="green seedling leaf"
                     /lab_host="TJC121"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
```

XhoI; Morex (mla) plants were greenhouse grown in the R Wise lab at Iowa State University, Ames, IA; 7 day old green seedlings were infected with isolate 5874 of Blumeria graminis f. sp. hordei, and leaves were harvested 24, 48 and 72 hr post-inoculation and snap frozen (Wise). In the TJ Close lab at the University of California, Riverside, total RNA was prepared from each sample pool, equal quantities of all three RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids (Chin). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 74 a 142 c 556 g 46 t
ORIGIN

Query Match 4.1%; Score 54.2; DB 10; Length 818;
Best Local Similarity 49.8%; Pred. No. 0.34;
Matches 137; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

```
Qy    814 CCCGTGCCCACTCTCCACGTCCACGGCACCATCCCTCTGCAGCCGCTCACCAGCCATGCC 873
          |||   ||| | | || |  || | || ||  ||| |  |  || |   ||  ||   ||
Db    512 CCCCCCCCCCCCCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCCCCACCCCCCCTCTCC 453

Qy    874 GTGTCGCGGAACGGCACAACCACCCCCAACCCACTCACGAAACCCCGTCCCGGCCGTGCC 933
          | |    |  | | || |||||  ||| | | |    ||||  |||  ||   ||
Db    452 CCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCC 393

Qy    934 CGTGTCGGTCCGCGCTCGGCAACGAGGCGGCCCGCGCTGCTGAGTCCCCTGGACACCCGA 993
          |    |   |||| | |    |||   | |||| | |  |  | |||||    | |||
Db    392 CCCCCCTTCCCGCCCCCCCTCACGCCCCCGCCCCCCCCCCCCCACTCCCCCCCCCCCCCCC 333

Qy    994 CACCCTGTCGGCCCTTTGTTTATTCATCCCGAAATCTCATCTGCCCCCACGGCCGACTGC 1053
          |||||   |  |||       ||||  |||   | |   ||  ||| | | ||| ||
Db    332 CACCCCCCCCCCCCCCCCCCGCATTCTCCCCCCCCAGCCACTCGCCCCGCTCCGCCCACGCG 273

Qy   1054 GCTGCGCCGCCCGGATATATATACCCATCGTTATC 1088
           | || ||  ||| | |       || ||   |||
Db    272 CCCGCCCCCGCCGCACAGCCCCCGCCTTCTCCATC 238
```

RESULT 12
AQ251130
LOCUS AQ251130 798 bp DNA linear GSS 07-OCT-1998
DEFINITION F26H6-Sp6 IGF Arabidopsis thaliana genomic clone F26H6, DNA sequence.
ACCESSION AQ251130
VERSION AQ251130.1 GI:3704196
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 798)
AUTHORS Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and Ecker,J.
TITLE BAC End Sequences at ATGC
JOURNAL Unpublished (1997)
COMMENT Contact: Ecker J.
Arabidopsis Thaliana Genome Center
University of Pennsylvania
Dept. of Biology, University of Pennsylvania, Philadelphia, PA 19104
Tel: 215-898-9384
Fax: 215-898-8780
Email: jecker@atgenome.bio.upenn.edu
Seq primer: Sp6
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .798
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="F26H6"
/clone_lib="IGF"
/sex="hermaphrodite"
/note="Vector: BeloBACII; Site_1: EcoRI; Site_2: EcoRI; Produced by Thomas Altmann"
BASE COUNT 347 a 99 c 90 g 259 t 3 others
ORIGIN

Query Match 4.1%; Score 54; DB 12; Length 798;
Best Local Similarity 53.9%; Pred. No. 0.37;
Matches 132; Conservative 0; Mismatches 111; Indels 2; Gaps 1;

```
Qy    226 TGCATATTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACATGCACATACAT 285
          ||  |||| | || ||  |  |      |  |  | ||||  |  | | |   |     |
Db    152 TGATTATTGTAATAGTAATCAAAAAGGTGAATTAATAGTATTTAACATAATAATTAGTAA 211

Qy    286 ATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTT 345
          |  ||||  ||| | |   |  |||  ||| || | |||||| |||| |    | |  |
Db    212 AAAATCATTTATTTGAATTTTGTTTATTAATTTCATTTAAAATTAAAGTTCTGCATAAAT 271

Qy    346 CTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTA 405
          |||  |  ||  | ||| |||  || ||| ||| |  ||||  |||| || | |     |
Db    272 NTAAAATAAATTATATTATAAATTTCTCTTCAAAACATTTATTTATTTTATAATAAAAAA 331

Qy    406 TTTCGAATTTCACTCTATAAACAACATAGTCTACAATGGAAAACAGTGCTTTGTACGACT 465
          ||   ||||  | |  | |   || |||   | || |  |||||| |  ||| ||| | |
Db    332 TT--AAATTATATTACAAATTTAAAATAAGTTTCACTTTAAAACATTTATTTATACTAAT 389

Qy    466 ATATA 470
           || |
Db    390 TTAAA 394
```

RESULT 13
CNS00FVE
LOCUS CNS00FVE 1101 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR32P18 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL071298
VERSION AL071298.1 GI:4951138
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and

Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR32P18"
                     /note="end : TET3"
BASE COUNT      267 a    223 c     31 g    184 t    396 others
ORIGIN
```

```
  Query Match             4.1%;  Score 54;  DB 12;  Length 1101;
  Best Local Similarity   29.1%;  Pred. No. 0.38;
  Matches  102;  Conservative   89;  Mismatches  159;  Indels    0;  Gaps    0;

Qy     204 TACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAG 263
           |: |:: : | |   |: :|: |  |::  :  : |      || |  | || | | :|
Db     551 TWAAWHAWWTMCAAATHTWTWATTAAWMMAYAAWATAAACCATAATTTTAATCCAAWWAM 610

Qy     264 TTATACATACATGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATT 323
           :| :: |::|:   :  |: ||  | | : ||:| : |  : |  |:  | :::|: ||:
Db     611 WTTHMWAWMCWCWAMACTWWATCCCTTTWAATHTTWTATTMAAAATWATTTWMWTWTATY 670

Qy     324 AAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACT 383
           |:|:   ::: |     |:::|   | |  :|  ::: | : |: |       |::  ||
Db     671 AHAWAATWMWTTCCWMTAMWWTTATAAAAWMAATWMWATAHTAWATAMATCTTAWYTTCT 730

Qy     384 TTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAACAACATAGTCTACAATG 443
             |  || || :| ||| |: :  ||  |   | | | : :||:  ::|:: :  :  ::
Db     731 ATCTCTTTTAYAWATTTTYTYYATTTCACAATAAWTTWCWAAMWTMMAWWTWTAMACMWW 790

Qy     444 GAAAACAGTGCTTTGTACGACTATATACGCGATGTGTGGCTACAACATAAGACAATATAG 503
           ||| |  |  : |  |  ::  ::::  | :| | :   |  : :|: : ::| :::
Db     791 CAAATCTATTHAATYAATCWYATYWWHWWCTWTWTTHTHWTTTHTHAYCHTHMACWHWCW 850

Qy     504 TCGTTTGAAGATTGAACCTATATATCGGTACGGTTAATCCGTCTATGTAC 553
           || ||: || | |  |  |:::|  |  | :  :|:|:|: | :|| : |
Db     851 TCHTTYCAAWACTWTAAATMWWTTCCCCTCMTWWTHAYCMTTTWATCWTC 900
```

```
RESULT 14
CNS008WC
LOCUS       CNS008WC                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR18L14 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL052719
VERSION     AL052719.1  GI:4934268
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
```

melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

```
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR18L14"
                     /note="end : TET3"
BASE COUNT      145 a    171 c    122 g    340 t    323 others
ORIGIN
```

```
  Query Match             4.1%;  Score 53.4;  DB 12;  Length 1101;
  Best Local Similarity   37.9%;  Pred. No. 0.5;
  Matches   83;  Conservative   44;  Mismatches   92;  Indels    0;  Gaps    0;

Qy     215 CTTCTTATTTGTGCATATTTTTATTGTTGTTTACTGATTATACGAGTAGTTATACATACA 274
           | | || |   |   | ||||| ||:|| |  | :  ::  ||   :|  || | :|:::
Db     596 CGTTTTTTANTTTNTTTTTTTTTTTBTTTTCAATYHYYYCAACCYAWAWATAAAAWTWMM 655

Qy     275 TGCACATACATATCATCACATATATCACAATATTTTTCTAAATTAAATTAAAACTAAAAA 334
             |:|||:| |  |   :  |: ::|:||     :: ||||  | | :|:|   | ||
Db     656 AAAAYATAYAAACAAAAYYYMAWYYMAHAAAYCAAYWATAAAAAACAAWAWATAAATAAT 715

Qy     335 TGACTAAATTTCTAACACCAACGACATTGTAATGTTTTCTCCAACAACTTTACCTATTCT 394
           | |: :|||:||  |  :|:   | : |  : ||||:|     || |:| ::::: |
Db     716 AAAATMTHTTTYTATTAAAWAHATWAAWATTTHWTTTTYTTTTTWAAATYTTYYWWWWWT 775

Qy     395 ACATTGTTCTATTTCGAATTTCACTCTATAAACAACATA 433
           ||: ||   ::|  ::|:|:  |  ||| | :: |||
Db     776 TWATYTTTAATWWTTAWWTYTYCTTAAATATAAWWWATA 814
```

```
RESULT 15
CNS04D0K
LOCUS       CNS04D0K                945 bp    DNA     linear   GSS 21-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            101H21 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL285149
VERSION     AL285149.1  GI:8023560
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 945)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 945)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 945)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
```

```
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .945
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="101H21"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG101CD11LP1~end : T7"
BASE COUNT      386 a    112 c     96 g    231 t    120 others
ORIGIN

  Query Match             4.1%;  Score 53.2;  DB 12;  Length 945;
  Best Local Similarity   35.9%;  Pred. No. 0.55;
  Matches  115;  Conservative   44;  Mismatches  161;  Indels    0;  Gaps    0;

Qy     128 CTACATCGAACATATACACGAGATGTCTCGTGTGAATAGAGTCACTAATGCCTTAAGCAT 187
           :  :: | | ||   :       | | | :  :    | |  || | |   |  ::  :
Db     376 SNNBMCCTACCAATCRNNNNNNKTTTTTNVNHWTTTAAYAANCATTTAMNACAMMHTTWN 435

Qy     188 CGGTTACTCCGTAGGGTACATTCTGTTCTTCTTATTTGTGCATATTTTTATTGTTGTTTA 247
              ||  |   ||   |  ::: | || || ||||   |    |  :|||   :  |:|:
Db     436 TTTTTTTTTANTANYWTTAWWWTTWTTTTTTTTTATAATTAATAAAAWTTAAANWATTWTW 495

Qy     248 CTGATTATACGAGTAGTTATACATACATGCACATACATATCATCACATATATCACAATAT 307
           |: : |  |  |  || : :::  |  | | | |:|| || | || : : | |::||
Db     496 AAAAWAWAAATAAAAAATAAWAWWWTTTTTAAAAAAAWATAATAAWATTWTWTAWAWWAT 555

Qy     308 TTTTCTAAATTAAATTAAAACTAAAAATGACTAAATTTCTAACACCAACGACATTGTAAT 367
           |||: ||  ||  ||| ||:   || || |  || |||  || :  |      |  :|||
Db     556 TTTWATATTTTTATTTAAWAATAATATAAAAAATTTTAAAAAWAAATATTTTTAAWAAT 615

Qy     368 GTTTTCTCCAACAACTTTACCTATTCTACATTGTTCTATTTCGAATTTCACTCTATAAAC 427
              :     || || | ||  |||   | |:   : :::::  ::  : :    ||| |
Db     616 AAAWATAAAAATAAATATATTTATAWAAWAWATAWAWWWWWTAWWAAWWWTAWAATATAA 675

Qy     428 AACATAGTCTACAATGGAAA 447
           || |||   :| |||  |||
Db     676 AAAATAAAAWAWAATAAAAA 695


Search completed: November  7, 2002, 12:25:35
Job time : 1252.03 secs
```